MapReduce Service

User Guide

Date **2022-11-30**

Contents

1 Overview

1.1 What Is MRS?

Big data is a huge challenge facing the Internet era as the data volume and types increase rapidly. Conventional data processing technologies, such as single-node storage and relational databases, are unable to solve the emerging big data problems. In this case, the Apache Software Foundation (ASF) has launched an open source Hadoop big data processing solution. Hadoop is an open source distributed computing platform that can fully utilize computing and storage capabilities of clusters to process massive amounts of data. If enterprises deploy Hadoop systems by themselves, the disadvantages include high costs, long deployment period, difficult maintenance, and inflexible use.

To solve the preceding problems, the cloud provides MapReduce Service (MRS) for managing the Hadoop system. With MRS, you can deploy a Hadoop cluster in just one click. MRS provides enterprise-level big data clusters on the cloud. Tenants can fully control clusters and easily run big data components such as Storm, Hadoop, Spark, HBase, and Kafka. MRS is fully compatible with open source APIs, and incorporates advantages of the cloud computing and storage and big data industry experience to provide customers with a full-stack big data platform featuring high performance, low cost, flexibility, and ease-of-use. In addition, the platform can be customized based on service requirements to help enterprises quickly build a massive data processing system and discover new value points and business opportunities by analyzing and mining massive amounts of data in real time or in non-real time.

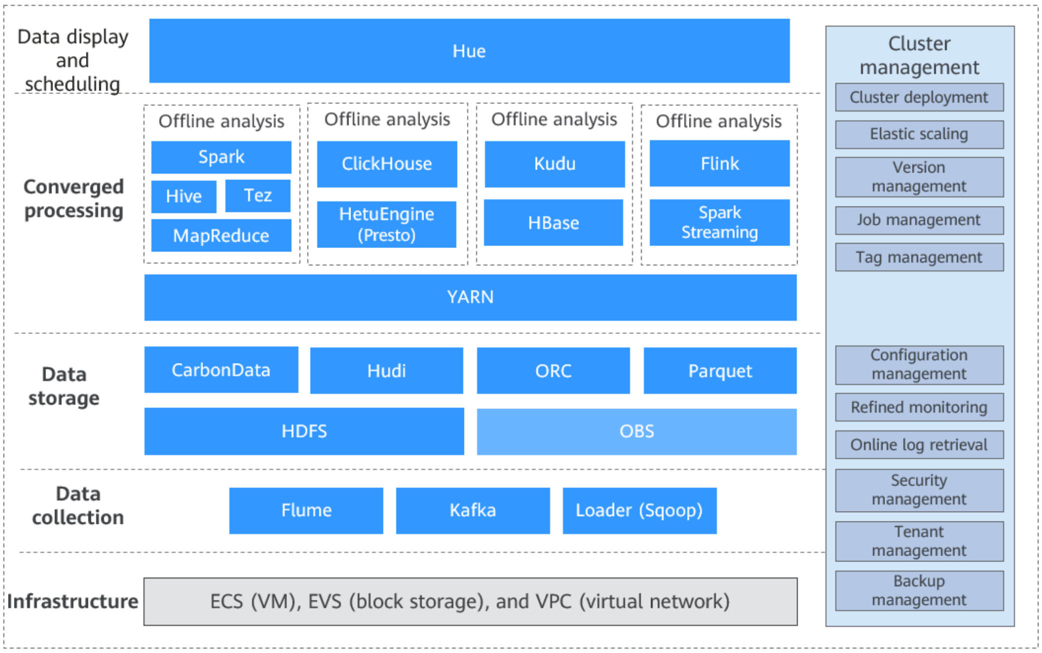
Product Architecture

Figure 1-1 shows the MRS logical architecture.

NOTE

MRS 3.x or later does not support patch management on the management console.

Figure 1-1 MRS architecture



MRS architecture includes infrastructure and big data processing phases.

- Infrastructure

  MRS big data clusters are built based on Elastic Cloud Server (ECS), and make full use of the high reliability and security capabilities of the virtualization layer.

  – A Virtual Private Cloud (VPC) is a virtual internal network provided for each tenant. It is isolated from other networks by default.

  – Elastic Volume Service (EVS) provides highly reliable and high-performance storage.

  – ECS provides scalable VMs, and works with VPCs, security groups, and the EVS multi-replica mechanism to build an efficient, reliable, and secure computing environment.

- Data collection

  The data collection layer provides the capability of importing data from various dta sources, such as Flume (data ingestion), Loader (relational data import), and Kafka (highly reliable message queue), to MRS big data clusters. Alternatively, you can use Cloud Data Migration (CDM) to import external data to MRS clusters.

- Data storage

  MRS clusters can store structured and unstructured data, and support multiple efficient formats to meet the requirements of different computing engines.

  – HDFS is a general-purpose distributed file system on a big data platform.

  – OBS is an object storage service that features high availability and low cost.

  – HBase supports data storage with indexes, and is applicable to high-performance index-based query scenarios.

- Data convergence processing
  - MRS provides multiple mainstream compute engines, including MapReduce (batch processing), Tez (DAG model), Spark (in-memory computing), Spark Streaming (micro-batch stream computing), Storm (stream computing), and Flink (stream computing), to convert data structures and logic into data models that meet service requirements in a variety of big data application scenarios.
  - Based on the preset data model and easy-to-use SQL data analysis, users can select Hive (data warehouse), SparkSQL, and Presto (interactive query engine).
- Data display and scheduling

  Displays data analysis results and integrates with Data Lake Governance Center (DGC) to provide a one-stop big data collaborative development platform, helping you easily complete multiple tasks, such as data modeling, data integration, script development, job scheduling, and O&M monitoring, making big data more accessible than ever before, and helping you effortlessly build big data processing centers.
- Cluster management

  All components of the Hadoop-based big data ecosystem are deployed in distributed mode, and their deployment, management, and O&M are complex.

  MRS provides a unified O&M management platform for cluster management, supporting one-click cluster deployment, multi-version selection, as well as manual scaling and auto scaling of clusters without service interruption. In addition, MRS provides job management, resource tag management, and O&M of the preceding data processing components at each layer. It also provides one-stop O&M capabilities, covering monitoring, alarm reporting, configuration, and patch upgrade.

Product Advantages

MRS has a powerful Hadoop kernel team and is deployed based on enterprise-level FusionInsight big data platform. MRS has been deployed on tens of thousands of nodes and can ensure Service Level Agreements (SLAs) for multi-level users.

MRS has the following advantages:

- High performance

  MRS supports self-developed CarbonData storage technology. CarbonData is a high-performance big data storage solution. It allows one data set to apply to multiple scenarios and supports features, such as multi-level indexing, dictionary encoding, pre-aggregation, dynamic partitioning, and quasi-real-time data query. This improves I/O scanning and computing performance and returns analysis results of tens of billions of data records in seconds. In addition, MRS supports self-developed enhanced scheduler Superior, which breaks the scale bottleneck of a single cluster and is capable of scheduling over 10,000 nodes in a cluster.
- Cost-effectiveness

  Based on diversified cloud infrastructure, MRS provides various computing and storage choices and separates computing from storage, delivering cost-effective massive data storage solutions. MRS supports auto scaling to address peak and off-peak service loads, releasing idle resources on the big data platform for customers. MRS clusters can be created and scaled out when you need them, and can be terminated or scaled in after you use them, minimizing cost.

- High security

  MRS delivers enterprise-level big data multi-tenant permissions management and security management to support table-based and column-based access control and data encryption.

- Easy O&M

  MRS provides a visualized big data cluster management platform, improving O&M efficiency. MRS supports rolling patch upgrade and provides visualized patch release information and one-click patch installation without manual intervention, ensuring long-term stability of user clusters.

- High reliability

  The proven large-scale reliability and long-term stability of MRS meet enterprise-level high reliability requirements. In addition, MRS supports automatic data backup across AZs and regions, as well as automatic anti-affinity. It allows VMs to be distributed on different physical machines.

1.2 Application Scenarios

Big data is ubiquitous in people's lives. MRS is suitable to process big data in the industries such as the Internet of things (IoT), e-commerce, finance, manufacturing, healthcare, energy, and government departments.

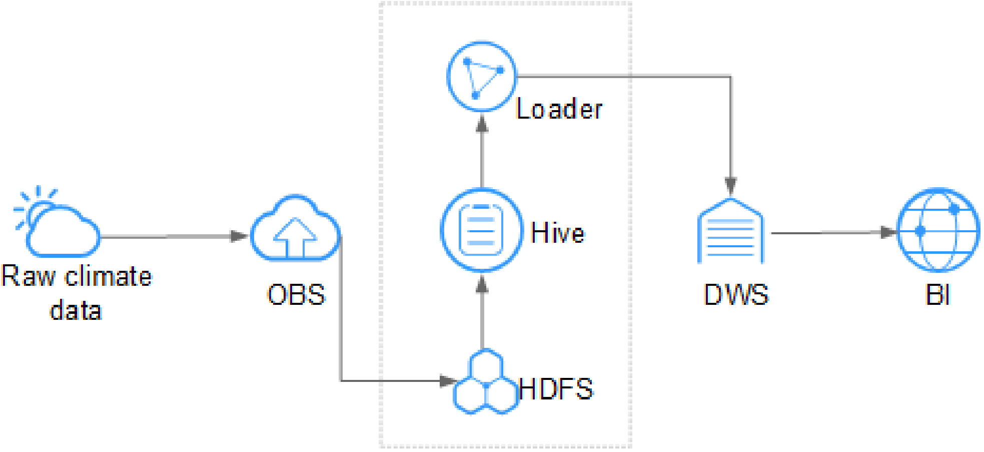
Large-scale data analysis

Large-scale data analysis is a major scenario in modern big data systems. Generally, an enterprise has multiple data sources. After data is accessed,extract, transform, and load (ETL) processing is required to generate modelized data for each service module to analyze and sort out data. This type of service has the following characteristics:

- The requirements for real-time execution are not high, and job execution time ranges from dozens of minutes to hours.
- The data volume is large.
- There are various data sources and diversified formats.
- Data processing usually consists of multiple tasks, and resources need to be planned in detail.

In the environmental protection industry, climate data is stored on OBS and periodically dumped into HDFS for batch analysis. 10 TB of climate data can be analyzed in 1 hour.

Figure 1-2 Large-scale data analysis in the environmental protection industry



MRS has the following advantages in this scenario.

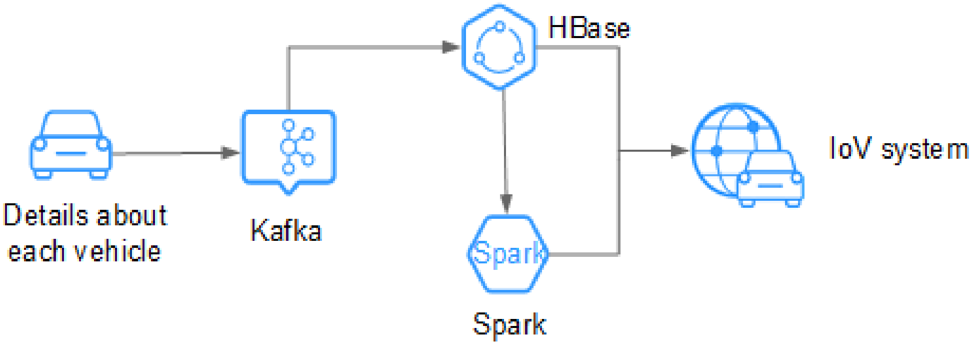
- Low cost: OBS offers cost-effective storage.
- Massive data analysis: TB/PB-level data is analyzed by Hive.
- Visualized data import and export tool: Loader exports data to Data Warehouse Service (DWS) for business intelligence (BI) analysis.

Large-scale data storage

A user who has a large amount of structured data usually requires index-based quasi-real-time query capabilities. For example, in an Internet of Vehicles (IoV) scenario, vehicle maintenance information is queried by vehicle number. Therefore, vehicle information is indexed based on vehicle numbers when it is being stored, to implement second-level response in this scenario. Generally, the data volume is large. The user may store data for one to three years.

For example, in the IoV industry, an automobile company stores data on HBase, which supports PB-level storage and CDR queries in milliseconds.

Figure 1-3 Large-scale data storage in the IoV industry



MRS has the following advantages in this scenario.

- Real time: Kafka accesses massive amounts of vehicle messages in real time.
- Massive data storage: HBase stores massive volumes of data and supports data queries in milliseconds.

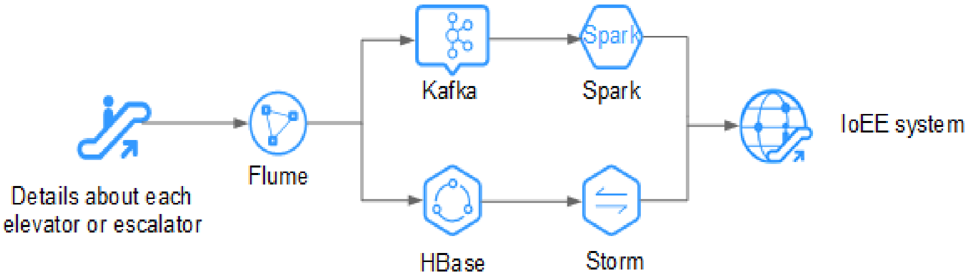
- Distributed data query: Spark analyzes and queries massive volumes of data.

Real-time data processing

Real-time data processing is usually used in scenarios such as anomaly detection, fraud detection, rule-based alarming, and service process monitoring. Data is processed while it is being inputted to the system.

For example, in the Internet of elevators & escalators (IoEE) industry, data of smart elevators and escalators is imported to MRS streaming clusters in real time for real-time alarming.

Figure 1-4 Low-latency streaming processing in the IoEE industry



MRS has the following advantages in this scenario.

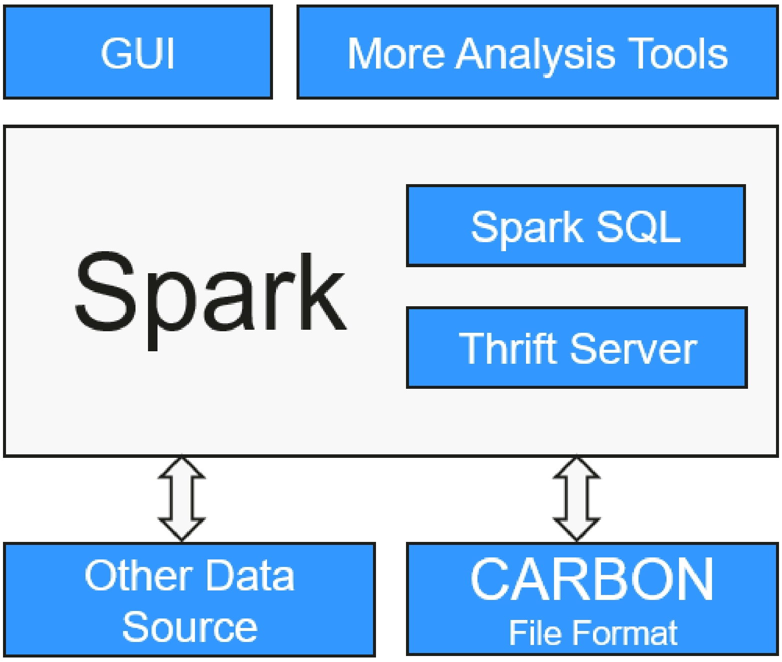
- Real-time data ingestion: Flume implements real-time data ingestion and provides various data collection and storage access methods.
- Data source access: Kafka accesses data of tens of thousands of elevators and escalators in real time.

1.3 Components

1.3.1 CarbonData

CarbonData is a new Apache Hadoop native data-store format. CarbonData allows faster interactive queries over PetaBytes of data using advanced columnar storage, index, compression, and encoding techniques to improve computing efficiency. In addition, CarbonData is also a high-performance analysis engine that integrates data sources with Spark.

Figure 1-5 Basic architecture of CarbonData



The purpose of using CarbonData is to provide quick response to ad hoc queries of big data. Essentially, CarbonData is an Online Analytical Processing (OLAP) engine, which stores data using tables similar to those in Relational Database Management System (RDBMS). You can import more than 10 TB data to tables created in CarbonData format, and CarbonData automatically organizes and stores data using the compressed multi-dimensional indexes. After data is loaded to CarbonData, CarbonData responds to ad hoc queries in seconds.

CarbonData integrates data sources into the Spark ecosystem. You can use Spark SQL to query and analyze data, or use the third-party tool ThriftServer provided by Spark to connect to Spark SQL.

CarbonData features

- SQL: CarbonData is compatible with Spark SQL and supports SQL query operations performed on Spark SQL.
- Simple Table dataset definition: CarbonData allows you to define and create datasets by using user-friendly Data Definition Language (DDL) statements. CarbonData DDL is flexible and easy to use, and can define complex tables.
- Easy data management: CarbonData provides various data management functions for data loading and maintenance. It can load historical data and incrementally load new data. The loaded data can be deleted according to the loading time and specific data loading operations can be canceled.
- CarbonData file format is a columnar store in HDFS. It has many features that a modern columnar format has, such as splittable and compression schema.

Unique features of CarbonData

- Stores data along with index: Significantly accelerates query performance and reduces the I/O scans and CPU resources, when there are filters in the query. CarbonData index consists of multiple levels of indices. A processing framework can leverage this index to reduce the task it needs to schedule and process, and it can also perform skip scan in more finer grain unit (called blocklet) in task side scanning instead of scanning the whole file.

- Operable encoded data: Through supporting efficient compression and global encoding schemes, CarbonData can query on compressed/encoded data. The data can be converted just before returning the results to the users, which is "late materialized".
- Supports various use cases with one single data format: like interactive OLAP-style query, Sequential Access (big scan), and Random Access (narrow scan).

Key technologies and advantages of CarbonData

- Quick query response: CarbonData features high-performance query. The query speed of CarbonData is 10 times of that of Spark SQL. It uses dedicated data formats and applies multiple index technologies, global dictionary code, and multiple push-down optimizations, providing quick response to TB-level data queries.
- Efficient data compression: CarbonData compresses data by combining the lightweight and heavyweight compression algorithms. This significantly saves 60% to 80% data storage space and the hardware storage cost.

For details about CarbonData architecture and principles, see **https://carbondata.apache.org/**.

1.3.2 ClickHouse

Introduction to ClickHouse

ClickHouse is an open-source columnar database oriented to online analysis and processing. It is independent of the Hadoop big data system and features ultimate compression rate and fast query performance. In addition, ClickHouse supports SQL query and provides good query performance, especially the aggregation analysis and query performance based on large and wide tables. The query speed is one order of magnitude faster than that of other analytical databases.

The core functions of ClickHouse are as follows:

Comprehensive DBMS functions

ClickHouse has comprehensive database management functions, including the basic functions of a Database Management System (DBMS):

- Data Definition Language (DDL): allows databases, tables, and views to be dynamically created, modified, or deleted without restarting services.
- Data Manipulation Language (DML): allows data to be queried, inserted, modified, or deleted dynamically.
- Permission control: supports user-based database or table operation permission settings to ensure data security.
- Data backup and restoration: supports data backup, export, import, and restoration to meet the requirements of the production environment.
- Distributed management: provides the cluster mode to automatically manage multiple database nodes.

Column-based storage and data compression

ClickHouse is a database that uses column-based storage. Data is organized by column. Data in the same column is stored together, and data in different columns is stored in different files.

During data query, columnar storage can reduce the data scanning range and data transmission size, thereby improving data query efficiency.

In a traditional row-based database system, data is stored in the sequence in **Table 1-1**:

Table 1-1 Row-based database

row	ID	Flag	Name	Event	Time
0	12345678901	0	name1	1	2020/1/11 15:19
1	32345678901	1	name2	1	2020/5/12 18:10
2	42345678901	1	name3	1	2020/6/13 17:38
N	...	...	...	...	...

In a row-based database, data in the same row is physically stored together. In a column-based database system, data is stored in the sequence in **Table 1-2**:

Table 1-2 Columnar database

row:	0	1	2	N
ID:	12345678901	32345678901	42345678901	...
Flag:	0	1	1	...
Name:	name1	name2	name3	...
Event:	1	1	1	...
Time:	2020/1/11 15:19	2020/5/12 18:10	2020/6/13 17:38	...

This example shows only the arrangement of data in a columnar database. Columnar databases store data in the same column together and data in different columns separately. Columnar databases are more suitable for online analytical processing (OLAP) scenarios.

Vectorized executor

ClickHouse uses CPU's Single Instruction Multiple Data (SIMD) to implement vectorized execution. SIMD is an implementation mode that uses a single instruction to operate multiple pieces of data and improves performance with data parallelism (other methods include instruction-level parallelism and thread-level parallelism). The principle of SIMD is to implement parallel data operations at the CPU register level.

Relational model and SQL query

ClickHouse uses SQL as the query language and provides standard SQL query APIs for existing third-party analysis visualization systems to easily integrate with ClickHouse.

In addition, ClickHouse uses a relational model. Therefore, the cost of migrating the system built on a traditional relational database or data warehouse to ClickHouse is lower.

Data sharding and distributed query

The ClickHouse cluster consists of one or more shards, and each shard corresponds to one ClickHouse service node. The maximum number of shards depends on the number of nodes (one shard corresponds to only one service node).

ClickHouse introduces the concepts of local table and distributed table. A local table is equivalent to a data shard. A distributed table itself does not store any data. It is an access proxy of the local table and functions as the sharding middleware. With the help of distributed tables, multiple data shards can be accessed by using the proxy, thereby implementing distributed query.

ClickHouse Applications

ClickHouse is short for Click Stream and Data Warehouse. It is initially applied to a web traffic analysis tool to perform OLAP analysis for data warehouses based on page click event flows. Currently, ClickHouse is widely used in Internet advertising, app and web traffic analysis, telecommunications, finance, and Internet of Things (IoT) fields. It is applicable to business intelligence application scenarios and has a large number of applications and practices worldwide. For details, visit **https://clickhouse.tech/docs/en/introduction/adopters/**.

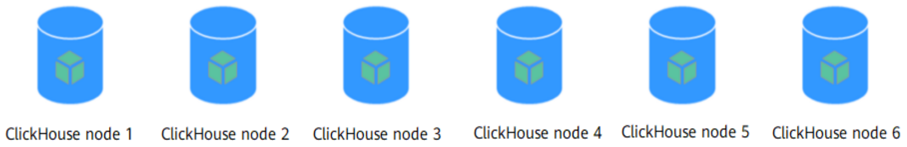
ClickHouse Enhanced Open Source Features

MRS ClickHouse has advantages such as automatic cluster mode, HA deployment, and smooth and elastic scaling.

- Automatic Cluster Mode

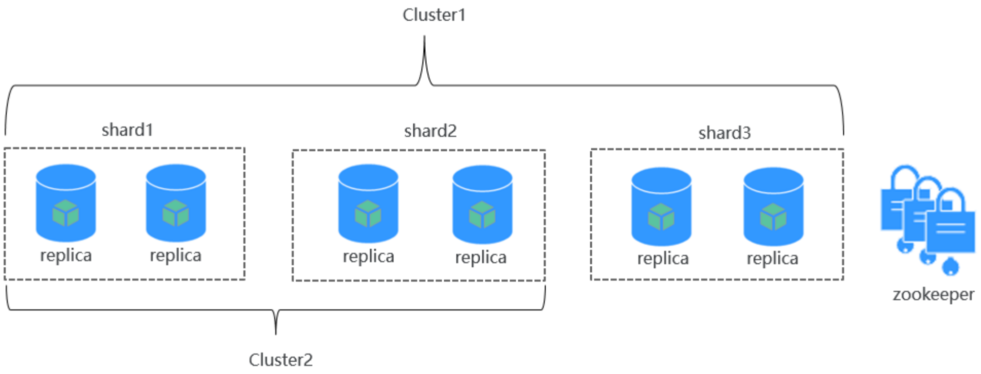
  As shown in **Figure 1-6**, a cluster consists of multiple ClickHouse nodes, which has no central node. It is more of a static resource pool. If the ClickHouse cluster mode is used for services, you need to pre-define the cluster information in the configuration file of each node. Only in this way, services can be correctly accessed.

  Figure 1-6 ClickHouse cluster



Users are unaware of data partitions and replica storage in common database systems. However, ClickHouse allows you to proactively plan and define detailed configurations such as shards, partitions, and replica locations. The ClickHouse instance of MRS packs the work in a unified manner and adapts it to the automatic mode, implementing unified management, which is flexible and easy to use. A ClickHouse instance consists of three ZooKeeper nodes and multiple ClickHouse nodes. The Dedicated Replica mode is used to ensure high reliability of dual data copies.

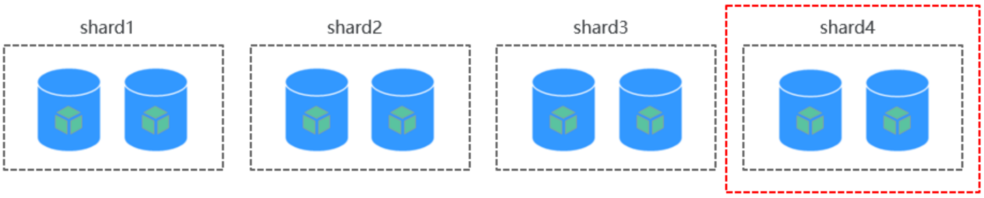
Figure 1-7 ClickHouse cluster structure



- Smooth and Elastic Scaling

  As business grows rapidly, MRS provides ClickHouse, a data migration tool, for scenarios such as the cluster's storage capacity or CPU compute resources approaching the limit. This tool is used to migrate some partitions of one or multiple MergeTree tables on several ClickHouseServer nodes to the same tables on other ClickHouseServer nodes. In this way, service availability is ensured and smooth capacity expansion is implemented.

  When you add ClickHouse nodes to a cluster, use this tool to migrate some data from the existing nodes to the new ones for data balancing after the expansion.

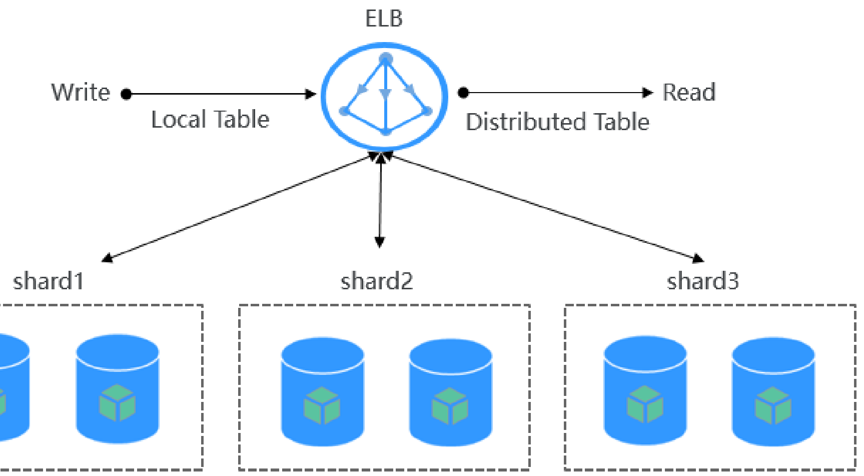


- HA Deployment Architecture

  MRS uses the ELB-based high availability (HA) deployment architecture to automatically distribute user access traffic to multiple backend nodes, expanding service capabilities to external systems and improving fault tolerance. As shown in **Figure 1-8**, when a client application requests a cluster, Elastic Load Balance (ELB) is used to distribute traffic. With the ELB polling mechanism, data is written to local tables and read from distributed tables on different nodes. In this way, data read/write load and high availability of application access are guaranteed.

  After the ClickHouse cluster is provisioned, each ClickHouse instance node in the cluster corresponds to a replica, and two replicas form a logical shard. For example, when creating a ReplicatedMergeTree table, you can specify shards so that data can be automatically synchronized between two replicas in the same shard.

Figure 1-8 HA deployment architecture



1.3.3 DBService

1.3.3.1 DBService Basic Principles

Overview

DBService is a HA storage system for relational databases, which is applicable to the scenario where a small amount of data (about 10 GB) needs to be stored, for example, component metadata. DBService can only be used by internal components of a cluster and provides data storage, query, and deletion functions.

DBService is a basic component of a cluster. Components such as Hive, Hue, Oozie, Loader, and Redis, and Loader store their metadata in DBService, and provide the metadata backup and restoration functions by using DBService.

DBService Architecture

DBService in the cluster works in active/standby mode. Two DBServer instances are deployed and each instance contains three modules: HA, Database, and FloatIP.

Figure 1-9 shows the DBService logical architecture.

Figure 1-9 DBService architecture

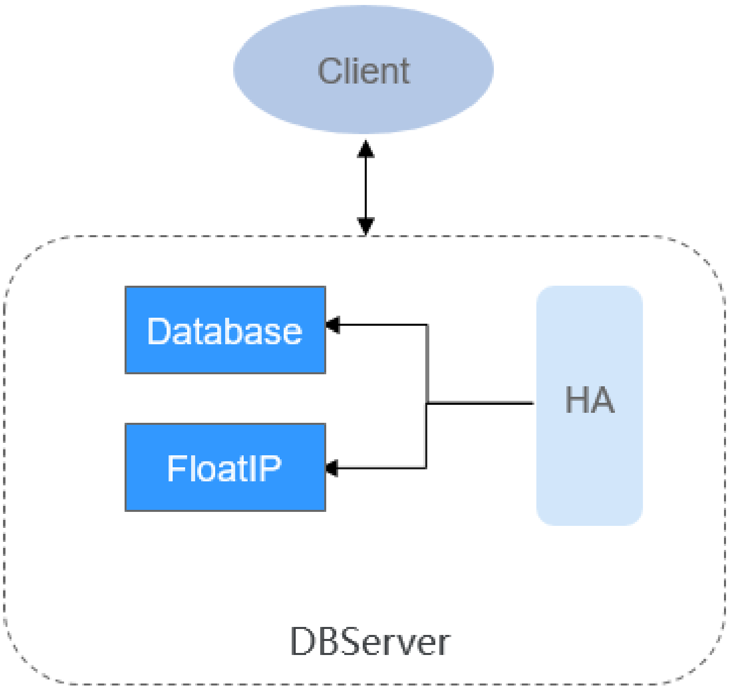


Table 1-3 describes the modules shown in **Figure 1-9**

Table 1-3 Module description

Name	Description
HA	HA management module. The active/standby DBServer uses the HA module for management.
Database	Database module. This module stores the metadata of the Client module.
FloatIP	Floating IP address that provides the access function externally. It is enabled only on the active DBServer instance and is used by the Client module to access Database.
Client	Client using the DBService component, which is deployed on the component instance node. The client connects to the database by using FloatIP and then performs metadata adding, deleting, and modifying operations.

1.3.3.2 Relationship Between DBService and Other Components

DBService is a basic component of a cluster. Components such as Hive, Hue, Oozie, Loader, Metadata, and Redis, and Loader store their metadata in DBService, and provide the metadata backup and restoration functions by using DBService.

1.3.4 Flink

1.3.4.1 Flink Basic Principles

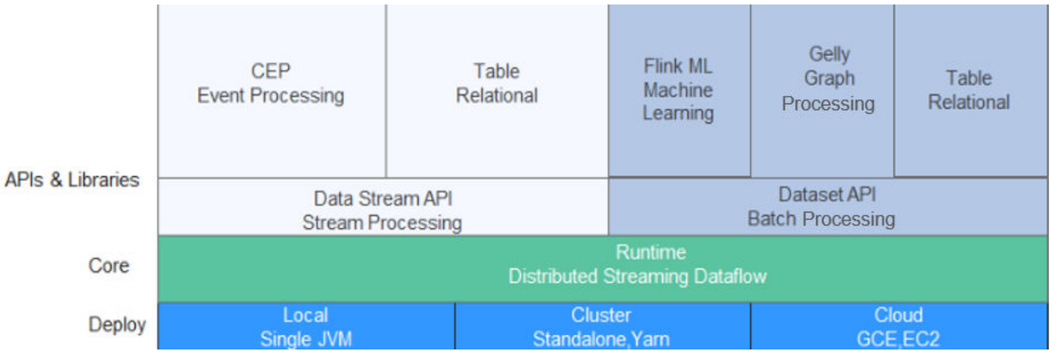
Overview

Flink is a unified computing framework that supports both batch processing and stream processing. It provides a stream data processing engine that supports data distribution and parallel computing. Flink features stream processing and is a top open source stream processing engine in the industry.

Flink provides high-concurrency pipeline data processing, millisecond-level latency, and high reliability, making it extremely suitable for low-latency data processing.

Figure 1-10 shows the technology stack of Flink.

Figure 1-10 Technology stack of Flink



Flink provides the following features in the current version:

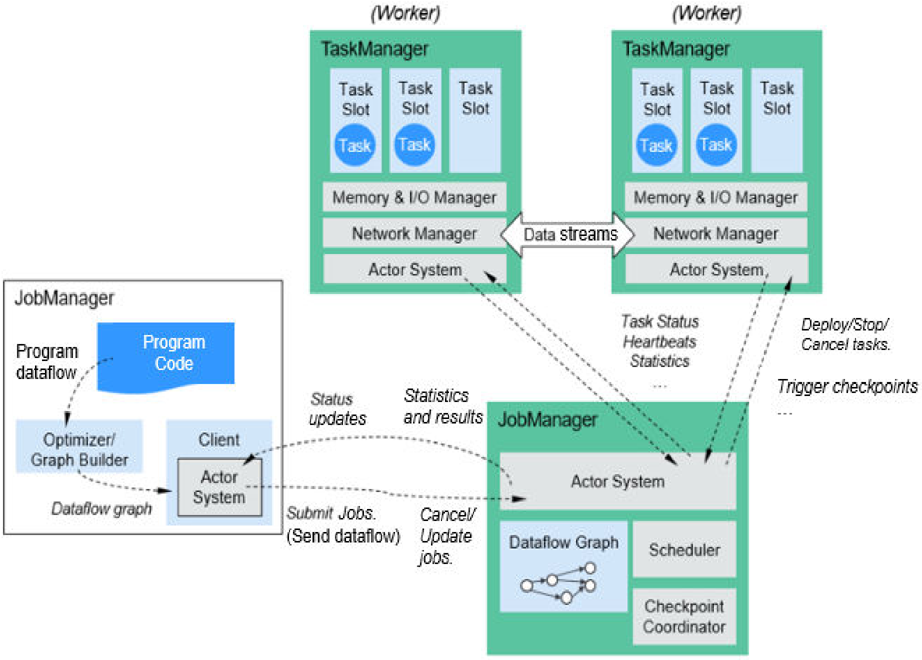
- DataStream
- Checkpoint
- Window
- Job Pipeline
- Configuration Table

Other features are inherited from the open source community and are not enhanced. For details, visit **https://ci.apache.org/projects/flink/flink-docs-release-1.12/**.

Flink Architecture

Figure 1-11 shows the Flink architecture.

Figure 1-11 Flink architecture



As shown in the above figure, the entire Flink system consists of three parts:

- Client

  Flink client is used to submit jobs (streaming jobs) to Flink.

- TaskManager

  TaskManager is a service execution node of Flink. It executes specific tasks. A Flink system can have multiple TaskManagers. These TaskManagers are equivalent to each other.

- JobManager

  JobManager is a management node of Flink. It manages all TaskManagers and schedules tasks submitted by users to specific TaskManagers. In high-availability (HA) mode, multiple JobManagers are deployed. Among these JobManagers, one is selected as the active JobManager, and the others are standby.

For more information about the Flink architecture, visit **https://ci.apache.org/projects/flink/flink-docs-master/docs/concepts/flink-architecture/**.

Flink Principles

- **Stream & Transformation & Operator**

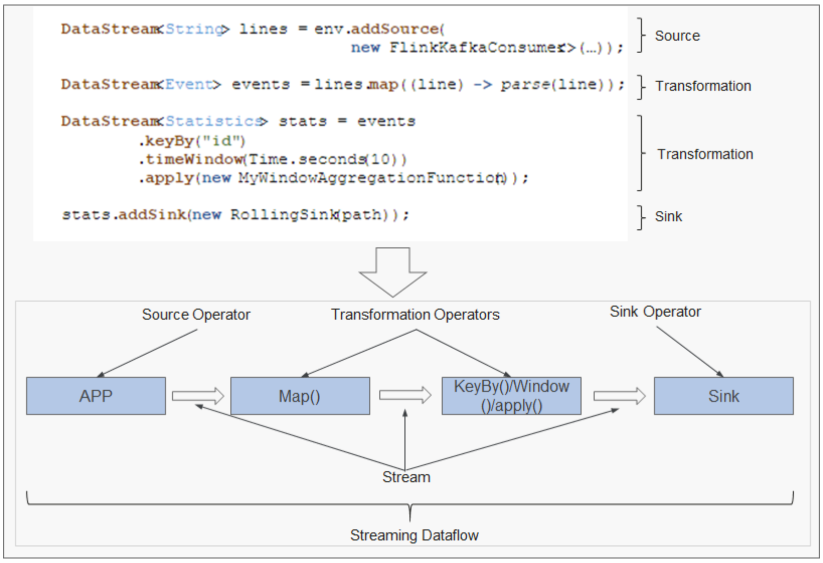
  A Flink program consists of two building blocks: stream and transformation.

  a. Conceptually, a stream is a (potentially never-ending) flow of data records, and a transformation is an operation that takes one or more streams as input, and produces one or more output streams as a result.

  b. When a Flink program is executed, it is mapped to a streaming dataflow. A streaming dataflow consists of a group of streams and transformation operators. Each dataflow starts with one or more source operators and ends in one or more sink operators. A dataflow resembles a directed acyclic graph (DAG).

Figure 1-12 shows the streaming dataflow to which a Flink program is mapped.

Figure 1-12 Example of Flink DataStream



```
DataStream<String> lines = env.addSource(
                                new FlinkKafkaConsumer<>(…));

DataStream<Event> events = lines.map((line) -> parse(line));

DataStream<Statistics> stats = events
        .keyBy("id")
        .timeWindow(Time.seconds(10))
        .apply(new MyWindowAggregationFunction());

stats.addSink(new RollingSink(path));
```

As shown in Figure 1-12, **FlinkKafkaConsumer** is a source operator; Map, KeyBy, TimeWindow, and Apply are transformation operators; RollingSink is a sink operator.

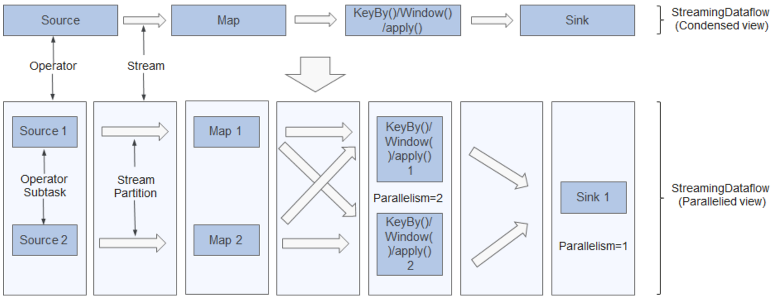
- **Pipeline Dataflow**

  Applications in Flink can be executed in parallel or distributed modes. A stream can be divided into one or more stream partitions, and an operator can be divided into multiple operator subtasks.

  The executor of streams and operators are automatically optimized based on the density of upstream and downstream operators.

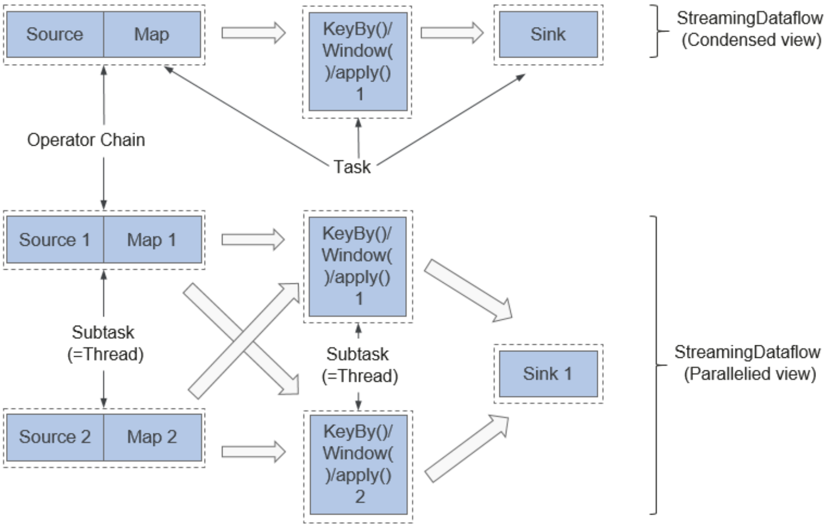
  – Operators with low density cannot be optimized. Each operator subtask is separately executed in different threads. The number of operator subtasks is the parallelism of that particular operator. The parallelism (the total number of partitions) of a stream is that of its producing operator. Different operators of the same program may have different levels of parallelism, as shown in Figure 1-13.

Figure 1-13 Operator



- Operators with high density can be optimized. Flink chains operator subtasks together into a task, that is, an operator chain. Each operator chain is executed by one thread on TaskManager, as shown in **Figure 1-14**.

Figure 1-14 Operator chain


Source
Map
KeyBy()/ Window( )/apply() 1
Sink
StreamingDataflow (Condensed view)
Operator Chain
Task
Source 1
Map 1
KeyBy()/ Window( )/apply() 1
Subtask (=Thread)
Subtask (=Thread)
Sink 1
StreamingDataflow (Parallelied view)
Source 2
Map 2
KeyBy()/ Window( )/apply() 2

- In the upper part of **Figure 1-14**, the condensed Source and Map operators are chained into an Operator Chain, that is, a larger operator. The Operator Chain, KeyBy, and Sink all represent an operator respectively and are connected with each other through streams. Each operator corresponds to one task during the running. Namely, there are three tasks in the upper part.
- In the lower part of **Figure 1-14**, each task, except Sink, is paralleled into two subtasks. The parallelism of the Sink operator is one.

Key Features

- Stream processing

  The real-time stream processing engine features high throughput, high performance, and low latency, which can provide processing capability within milliseconds.

- Various status management

  The stream processing application needs to store the received events or intermediate result in a certain period of time for subsequent access and processing at a certain time point. Flink provides diverse features for status management, including:

  - Multiple basic status types: Flink provides various states for data structures, such as ValueState, ListState, and MapState. Users can select the most efficient and suitable status type based on the service model.
  - Rich State Backend: State Backend manages the status of applications and performs Checkpoint operations as required. Flink provides different

  State Backends. State can be stored in the memory or RocksDB, and supports the asynchronous and incremental Checkpoint mechanism.
  - Exactly-once state consistency: The Checkpoint and fault recovery capabilities of Flink ensure that the application status of tasks is consistent before and after a fault occurs. Flink supports transactional output for some specific storage devices. In this way, exactly-once output can be ensured even when a fault occurs.

- Various time semantics

  Time is an important part of stream processing applications. For real-time stream processing applications, operations such as window aggregation, detection, and matching based on time semantics are very common. Flink provides various time semantics.

  - Event-time: The timestamp provided by the event is used for calculation, making it easier to process the events that arrive at a random sequence or arrive late.
  - Watermark: Flink introduces the concept of Watermark to measure the development of event time. Watermark also provides flexible assurance for balancing processing latency and data integrity. When processing event streams with Watermark, Flink provides multiple processing options if data arrives after the calculation, for example, redirecting data (side output) or updating the calculation result.
  - Processing-time and Ingestion-time are supported.
  - Highly flexible streaming window: Flink supports the time window, count window, session window, and data-driven customized window. You can customize the triggering conditions to implement the complex streaming calculation mode.

- Fault tolerance mechanism

  In a distributed system, if a single task or node breaks down or is faulty, the entire task may fail. Flink provides a task-level fault tolerance mechanism, which ensures that user data is not lost when an exception occurs in a task and can be automatically restored.

  - Checkpoint: Flink implements fault tolerance based on checkpoint. Users can customize the checkpoint policy for the entire task. When a task fails, the task can be restored to the status of the latest checkpoint and data after the snapshot is resent from the data source.
  - Savepoint: A savepoint is a consistent snapshot of application status. The savepoint mechanism is similar to that of checkpoint. However, the savepoint mechanism needs to be manually triggered. The savepoint mechanism ensures that the status information of the current stream application is not lost during task upgrade or migration, facilitating task suspension and recovery at any time point.

- Flink SQL

  Table APIs and SQL use Apache Calcite to parse, verify, and optimize queries. Table APIs and SQL can be seamlessly integrated with DataStream and DataSet APIs, and support user-defined scalar functions, aggregation functions, and table value functions. The definition of applications such as data analysis and ETL is simplified. The following code example shows how to use Flink SQL statements to define a counting application that records session times.

```
SELECT userId, COUNT(*)
FROM clicks
GROUP BY SESSION(clicktime, INTERVAL '30' MINUTE), userId
```

For more information about Flink SQL, see **https://ci.apache.org/projects/flink/flink-docs-master/dev/table/sqlClient.html**.

- CEP in SQL

  Flink allows users to represent complex event processing (CEP) query results in SQL for pattern matching and evaluate event streams on Flink.

  CEP SQL is implemented through the **MATCH_RECOGNIZE** SQL syntax. The **MATCH_RECOGNIZE** clause is supported by Oracle SQL since Oracle Database 12c and is used to indicate event pattern matching in SQL. The following is an example of CEP SQL:

```
SELECT T.aid, T.bid, T.cid
FROM MyTable
    MATCH_RECOGNIZE (
      PARTITION BY userid
      ORDER BY proctime
      MEASURES
        A.id AS aid,
        B.id AS bid,
        C.id AS cid
      PATTERN (A B C)
      DEFINE
        A AS name = 'a',
        B AS name = 'b',
        C AS name = 'c'
    ) AS T
```

1.3.4.2 Flink HA Solution

Flink HA Solution

A Flink cluster has only one JobManager. This has the risks of single point of failures (SPOFs). There are three modes of Flink: Flink On Yarn, Flink Standalone, and Flink Local. Flink On Yarn and Flink Standalone modes are based on clusters and Flink Local mode is based on a single node. Flink On Yarn and Flink Standalone provide an HA mechanism. With such a mechanism, you can recover the JobManager from failures and thereby eliminate SPOF risks. This section describes the HA mechanism of the Flink On Yarn.

Flink supports the HA mode and job exception recovery that highly depend on ZooKeeper. If you want to enable the two functions, configure ZooKeeper in the **flink-conf.yaml** file in advance as follows:

```
high-availability: zookeeper
high-availability.zookeeper.quorum:  ZooKeeper IP address:2181
high-availability.storageDir: hdfs:///flink/recovery
```

Flink On Yarn

Flink JobManager and Yarn ApplicationMaster are in the same process. Yarn ResourceManager monitors ApplicationMaster. If ApplicationMaster is abnormal, Yarn restarts it and restores all JobManager metadata from HDFS. During the recovery, existing tasks cannot run and new tasks cannot be submitted. ZooKeeper stores JobManager metadata, such as information about jobs, to be used by the new JobManager. A TaskManager failure is listened and processed by the DeathWatch mechanism of Akka on JobManager. When a TaskManager fails, a container is requested again from Yarn and a TaskManager is created.

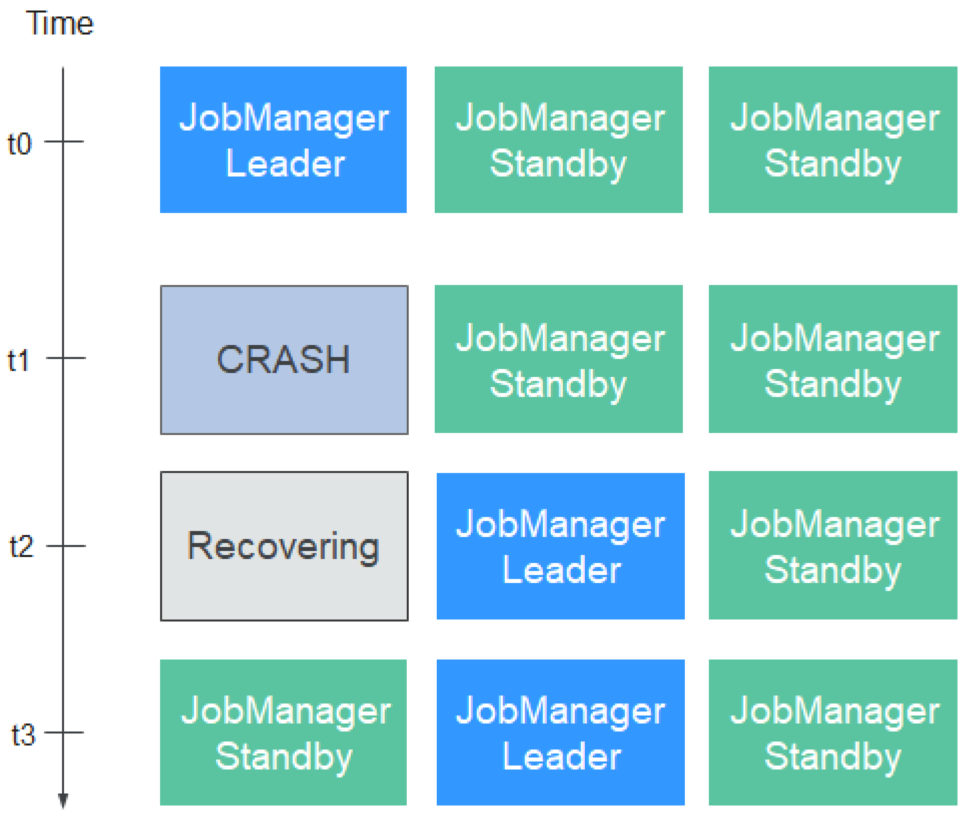
For more information about the HA solution of Flink on YARN, visit:

http://hadoop.apache.org/docs/r3.1.1/hadoop-yarn/hadoop-yarn-site/ResourceManagerHA.html

Standalone

In the standalone mode, multiple JobManagers can be started and ZooKeeper elects one as the leader JobManager. In this mode, there is a leader JobManager and multiple standby JobManagers. If the leader JobManager fails, a standby JobManager takes over the leadership. **Figure 1-15** shows the process of a leader/standby JobManager switchover.

Figure 1-15 Switchover process



Restoring TaskManager

A TaskManager failure is listened and processed by the DeathWatch mechanism of Akka on JobManager. If the TaskManager fails, the JobManager creates a TaskManager and migrates services to the created TaskManager.

Restoring JobManager

Flink JobManager and Yarn ApplicationMaster are in the same process. Yarn ResourceManager monitors ApplicationMaster. If ApplicationMaster is abnormal, Yarn restarts it and restores all JobManager metadata from HDFS. During the recovery, existing tasks cannot run and new tasks cannot be submitted.

Restoring Jobs

If you want to restore jobs, ensure that the startup policy is configured in Flink configuration files. Supported restart policies are **fixed-delay**, **failure-rate**, and **none**. Jobs can be restored only when the policy is configured to **fixed-delay** or **failure-rate**. If the restart policy is configured to **none** and checkpoint is configured for jobs, the restart policy is automatically configured to **fixed-delay** and the value of **restart-strategy.fixed-delay.attempts** (which specifies the number of retry times) is configured to **Integer.MAX_VALUE**.

For details about the three strategies, visit the Flink official website at **https://ci.apache.org/projects/flink/flink-docs-release-1.12/dev/task_failure_recovery.html**. The configuration strategies are as follows:

```
restart-strategy: fixed-delay
restart-strategy.fixed-delay.attempts: 3
restart-strategy.fixed-delay.delay: 10 s
```

Jobs will be restored in the following scenarios:

- If a JobManager fails, all its jobs are stopped, and will be recovered after another JobManager is created and running.
- If a TaskManager fails, all tasks on the TaskManager are stopped, and will be started until there are available resources.
- When a task of a job fails, the job is restarted.

NOTE

For details about how to configure job restart strategies, see **https://ci.apache.org/projects/flink/flink-docs-release-1.12/ops/jobmanager_high_availability.html**.

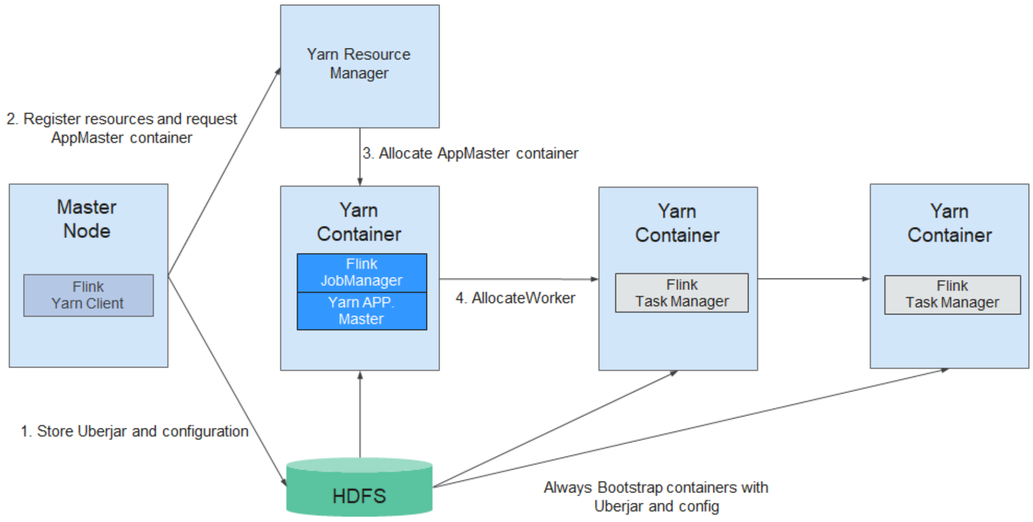
1.3.4.3 Relationship with Other Components

Relationship between Flink and Yarn

Flink supports Yarn-based cluster management mode. In this mode, Flink serves as an application of Yarn and runs on Yarn.

Figure 1-16 shows how Flink interacts with Yarn.

Figure 1-16 Flink interaction with Yarn



1. The Flink Yarn Client first checks whether there are sufficient resources for starting the Yarn cluster. If yes, the Flink Yarn client uploads JAR packages and configuration files to HDFS.
2. Flink Yarn client communicates with Yarn ResourceManager to request a container for starting ApplicationMaster. After all Yarn NodeManagers finish downloading the JAR package and configuration files, the ApplicationMaster is started.
3. During the startup, the ApplicationMaster interacts with the Yarn ResourceManager to request the container for starting a TaskManager. After the container is ready, the TaskManager process is started.
4. In the Flink Yarn cluster, the ApplicationMaster and Flink JobManager are running in the same container. The ApplicationMaster informs each TaskManager of the RPC address of the JobManager. After TaskManagers are started successfully, they register with the JobManager.
5. After all TaskManagers have registered with the JobManager successfully, Flink starts up in the Yarn cluster. Then, the Flink Yarn client can submit Flink jobs to the JobManager, and Flink can perform mapping, scheduling, and computing for the jobs.

1.3.4.4 Flink Enhanced Open Source Features

1.3.4.4.1 Window

Enhanced Open Source Feature: Window

This section describes the sliding window of Flink and provides the sliding window optimization method. For details about windows, visit the official website at **https://ci.apache.org/projects/flink/flink-docs-release-1.12/dev/stream/operators/windows.html**.

Introduction to Window

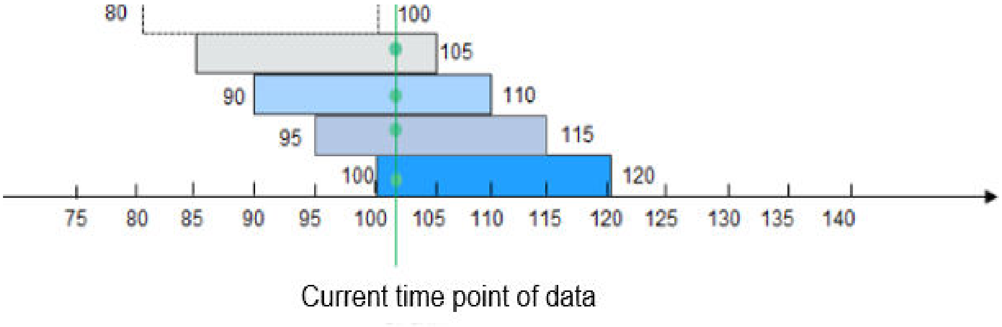
Data in a window is saved as intermediate results or original data. If you perform a sum operation (**window(SlidingEventTimeWindows.of(Time.seconds(20), Time.seconds(5))).sum**) on data in the window, only the intermediate result will be retained. If a custom window (**window(SlidingEventTimeWindows.of(Time.seconds(20), Time.seconds(5))).apply(new UDF)**) is used, all original data in the window will be saved.

If custom windows **SlidingEventTimeWindow** and **SlidingProcessingTimeWindow** are used, data is saved as multiple backups. Assume that the window is defined as follows:

```
window(SlidingEventTimeWindows.of(Time.seconds(20), Time.seconds(5))).apply(new
UDFWindowFunction)
```

If a block of data arrives, it is assigned to four different windows (20/5 = 4). That is, the data is saved as four copies in the memory. When the window size or sliding period is set to a large value, data will be saved as excessive copies, causing redundancy.

Figure 1-17 Original structure of a window



If a data block arrives at the 102nd second, it is assigned to windows [85, 105), [90, 110), [95, 115), and [100, 120).

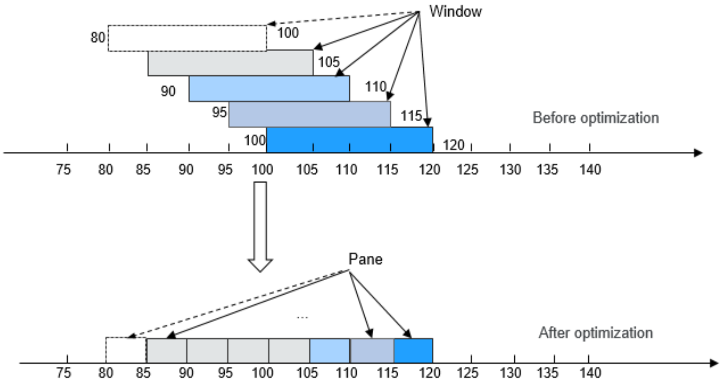
Window Optimization

As mentioned in the preceding, there are excessive data copies when original data is saved in SlidingEventTimeWindow and SlidingProcessingTimeWindow. To resolve this problem, the window that stores the original data is restructured, which optimizes the storage and greatly lowers the storage space. The window optimization scheme is as follows:

1. Use the sliding period as a unit to divide a window into different panes.

   A window consists of one or multiple panes. A pane is essentially a sliding period. For example, the sliding period (namely, the pane) of **window(SlidingEventTimeWindows.of(Time.seconds(20), Time.seconds.of(5)))** lasts for 5 seconds. If this window ranges from [100, 120), this window can be divided into panes [100, 105), [105, 110), [110, 115), and [115, 120).

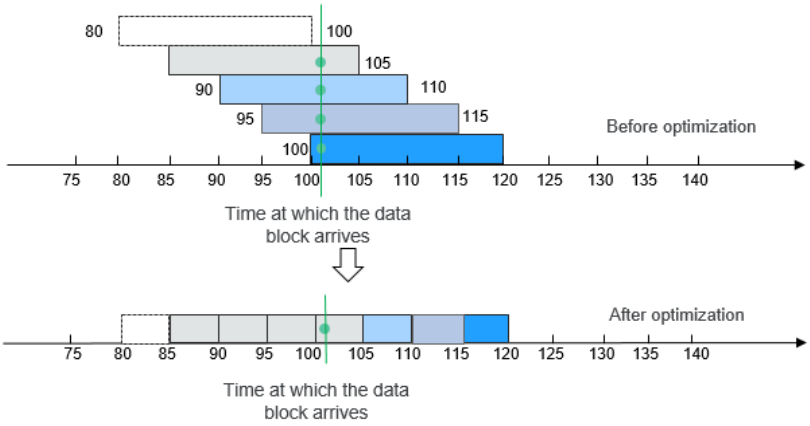
   Figure 1-18 Window optimization

2. When a data block arrives, it is not assigned to a specific window. Instead, Flink determines the pane to which the data block belongs based on the timestamp of the data block, and saves the data block into the pane.

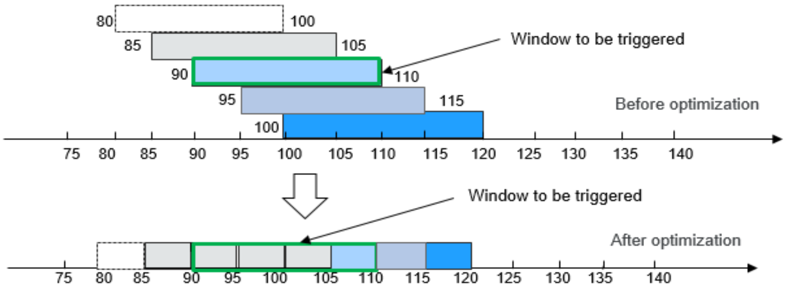
   A data block is saved only in one pane. In this case, only a data copy exists in the memory.

Figure 1-19 Saving data in a window



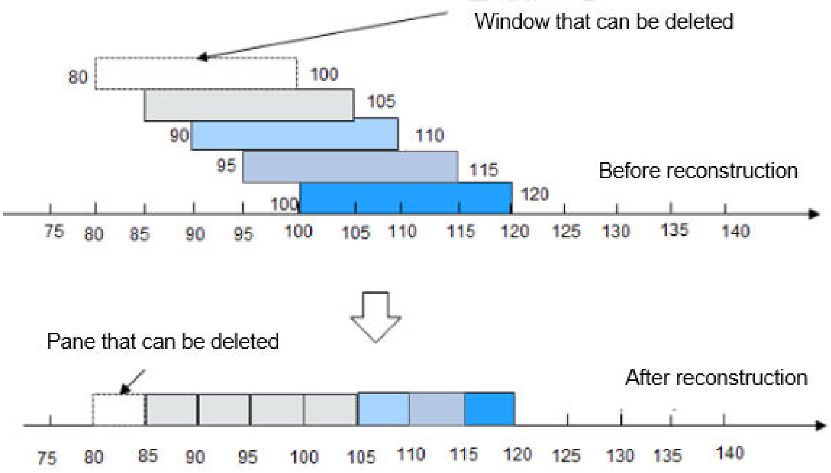
3. To trigger a window, compute all panes contained in the window, and combine all these panes into a complete window.

Figure 1-20 Triggering a window



4. If a pane is not required, you can delete it from the memory.

Figure 1-21 Deleting a window



After optimization, the quantity of data copies in the memory and snapshot is greatly reduced.

1.3.4.4.2 Job Pipeline

Enhanced Open Source Feature: Job Pipeline

Generally, logic code related to a service is stored in a large JAR package, which is called Fat JAR. Disadvantages of Fat JAR are as follows:

- When service logic becomes more and more complex, the size of the Fat JAR increases.
- Fat Jar makes coordination complex. Developers of all services are working with the same service logic. Even though the service logic can be divided into several modules, all modules are tightly coupled with each other. If the requirement needs to be changed, the entire flow diagram needs to be replanned.

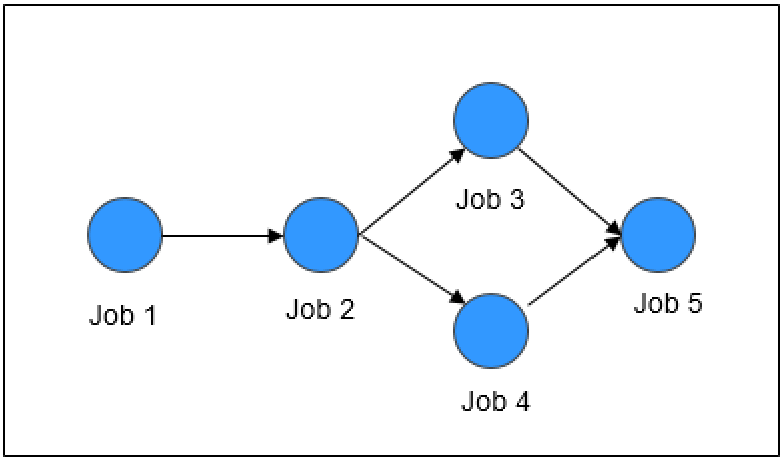
Splitting of jobs is facing the following problems:

- Data transmission between jobs can be achieved using Kafka. For example, job A transmits data to the topic A in Kafka, and then job B and job C read data from the topic A in Kafka. This solution is simple and easy to implement, but the latency is always longer than 100 ms.
- Operators are connected using the TCP protocol. In distributed environment, operators can be scheduled to any node and upstream and downstream services cannot detect the scheduling.

Job Pipeline

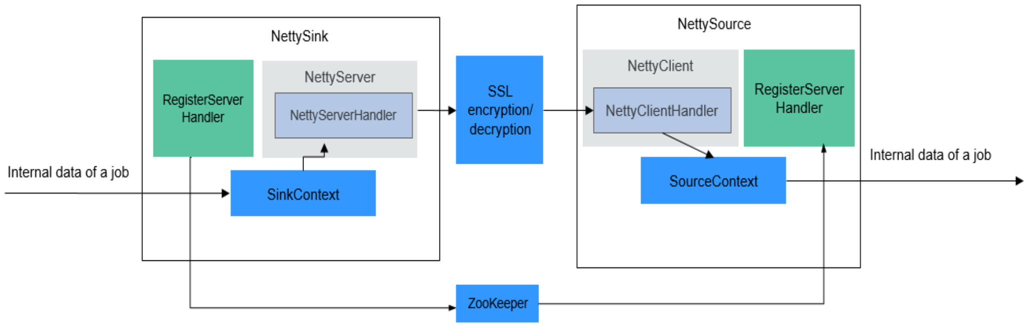
A pipeline consists of multiple Flink jobs connected through TCP. Upstream jobs can send data to downstream jobs. The flow diagram about data transmission is called a job pipeline, as shown in **Figure 1-22**.

Figure 1-22 Job pipeline



Job Pipeline Principles

Figure 1-23 Job pipeline principles



- NettySink and NettySource

  In a pipeline, upstream jobs and downstream jobs communicate with each other through Netty. The Sink operator of the upstream job works as a server and the Source operator of the downstream job works as a client. The Sink operator of the upstream job is called NettySink, and the Source operator of the downstream job is called NettySource.

- NettyServer and NettyClient

  NettySink functions as the server of Netty. In NettySink, NettyServer achieves the function of a server. NettySource functions as the client of Netty. In NettySource, NettyClient achieves the function of a client.

- Publisher

  The job that sends data to downstream jobs through NettySink is called a publisher.

- Subscriber

  The job that receives data from upstream jobs through NettySource is called a subscriber.

- RegisterServer

  RegisterServer is the third-party memory that stores the IP address, port number, and concurrency information about NettyServer.

- The general outside-in architecture is as follows:
  - NettySink->NettyServer->NettyServerHandler
  - NettySource->NettyClient->NettyClientHandler

Job Pipeline Functions

- **NettySink**

  NettySink consists of the following major modules:

  - RichParallelSinkFunction

    NettySink inherits RichParallelSinkFunction and attributes of Sink operators. The RichParallelSinkFunction API implements following functions:

    - Starts the NettySink operator.
    - Runs the NettySink operator and receives data from the upstream operator.

    - Cancels the running of NettySink operators.

  Following information can be obtained using the attribute of RichParallelSinkFunction:

    - subtaskIndex about the concurrency of each NettySink operator.
    - Concurrency of the NettySink operator.

- RegisterServerHandler

  RegisterServerHandler interacts with the component of RegisterServer and defines following APIs:

    - **start();**: Starts the RegisterServerHandler and establishes a contact with the third-party RegisterServer.
    - **createTopicNode();**: Creates a topic node.
    - **register();**: Registers information such as the IP address, port number, and concurrency to the topic node.
    - **deleteTopicNode();**: Deletes a topic node.
    - **unregister();**: Deletes registration information.
    - **query();**: Queries registration information.
    - **isExist();**: Verifies that a specific piece of information exists.
    - **shutdown();**: Disables the RegisterServerHandler and disconnects from the third-party RegisterServer.

  NOTE

  - RegisterServerHandler API enables ZooKeeper to work as the handler of RegisterServer. You can customize your handler as required. Information is stored in ZooKeeper in the following form:

    ```
    Namespace
    |---Topic-1
      |---parallel-1
      |---parallel-2
      |....
      |---parallel-n
    |---Topic-2
      |---parallel-1
      |---parallel-2
      |....
      |---parallel-m
    |...
    ```

  - Information about NameSpace can be obtained from the following parameters of the **flink-conf.yaml** file:

    ```
    nettyconnector.registerserver.topic.storage: /flink/nettyconnector
    ```

  - The simple authentication and security layer (SASL) authentication between ZookeeperRegisterServerHandler and ZooKeeper is implemented through the Flink framework.
  - Ensure that each job has a unique topic. Otherwise, the subscription relationship may be unclear.
  - When calling **shutdown()**, ZookeeperRegisterServerHandler deletes the registration information about the current concurrency, and then attempts to delete the topic node. If the topic node is not empty, deletion will be canceled, because not all concurrency has exited.

  - NettyServer

    NettyServer is the core of the NettySink operator, whose main function is to create a NettyServer and receive connection requests from NettyClient. Use NettyServerHandler to send data received from upstream operators of a same job. The port number and subnet of NettyServer needs to be configured in the **flink-conf.yaml** file.

    - Port range
      ```
      nettyconnector.sinkserver.port.range: 28444-28943
      ```
    - Subnet
      ```
      nettyconnector.sinkserver.subnet: 10.162.222.123/24
      ```

      NOTE

      The **nettyconnector.sinkserver.subnet** parameter is set to the subnet (service IP address) of the Flink client by default. If the client and TaskManager are not in the same subnet, an error may occur. Therefore, you need to manually set this parameter to the subnet (service IP address) of TaskManager.

  - NettyServerHandler

    The handler enables the interaction between NettySink and subscribers. After NettySink receives messages, the handler sends these messages out. To ensure data transmission security, this channel is encrypted using SSL. The **nettyconnector.ssl.enabled** configures whether to enable SSL encryption. The SSL encryption is enabled only when **nettyconnector.ssl.enabled** is set to **true**.

- **NettySource**

  NettySource consists of the following major modules:

  - RichParallelSourceFunction

    NettySource inherits RichParallelSinkFunction and attributes of Source operators. The RichParallelSourceFunction API implements following functions:

    - Starts the NettySink operator.
    - Runs the NettySink operator, receives data from subscribers, and injects the data to jobs.
    - Cancels the running of Source operators.

    Following information can be obtained using the attribute of RichParallelSourceFunction:

    - subtaskIndex about the concurrency of each NettySource operator.
    - Concurrency of the NettySource operator.

    When the NettySource operator enters the running stage, the NettyClient status is monitored. Once abnormality occurs, NettyClient is restarted and reconnected to NettyServer, preventing data confusion.

  - RegisterServerHandler

    RegisterServerHandler of NettySource has similar function as the RegisterServerHandler of NettySink. It obtains the IP address, port number, and information of concurrent operators of each subscribed job obtained in the NettySource operator.

- NettyClient

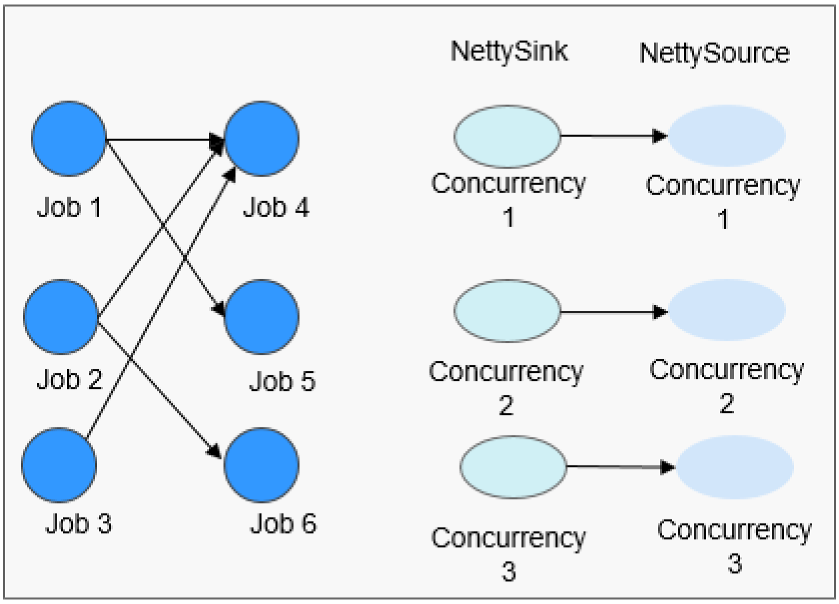
  NettyClient establishes a connection with NettyServer and uses NettyClientHandler to receive data. Each NettySource operator must have a unique name (specified by the user). NettyServer determines whether each client comes from different NettySources based on unique names. When a connection is established between NettyClient and NettyServer, NettyClient is registered with NettyServer and the NettySource name of NettyClient is transferred to NettyServer.

- NettyClientHandler

  The NettyClientHandler enables the interaction with publishers and other operators of the job. When messages are received, NettyClientHandler transfers these messages to the job. To ensure secure data transmission, SSL encryption is enabled for the communication with NettySink. The SSL encryption is enabled only when SSL is enabled and **nettyconnector.ssl.enabled** is set to **true**.

The relationship between the jobs may be many-to-many. The concurrency between each NettySink and NettySource operator is one-to-many, as shown in **Figure 1-24**.

Figure 1-24 Relationship diagram



1.3.4.4.3 Stream SQL Join

Enhanced Open Source Feature: Stream SQL Join

Flink's Table API&SQL is an integrated query API for Scala and Java that allows the composition of queries from relational operators such as selection, filter, and join in an intuitive way. For details about Table API & SQL, visit the official website at **https://ci.apache.org/projects/flink/flink-docs-release-1.12/dev/table/index.html**.

Introduction to Stream SQL Join

SQL Join is used to query data based on the relationship between columns in two or more tables. Flink Stream SQL Join allows you to join two streaming tables and query results from them. Queries similar to the following are supported:

```
SELECT o.proctime, o.productId, o.orderId, s.proctime AS shipTime
FROM Orders AS o
JOIN Shipments AS s
ON o.orderId = s.orderId
AND o.proctime BETWEEN s.proctime AND s.proctime + INTERVAL '1' HOUR;
```

Currently, Stream SQL Join needs to be performed within a specified window. The join operation for data within the window requires at least one equi-join predicate and a join condition that bounds the time on both sides. Such a condition can be defined by two appropriate range predicates (**<**, **<=**, **>=**, **>**), a **BETWEEN** predicate, or a single equality predicate that compares the same type of time attributes (such as processing time or event time) of both input tables.

The following example will join all orders with their corresponding shipments if the order was shipped four hours after the order was received.

```
SELECT *
FROM Orders o, Shipments s
WHERE o.id = s.orderId AND
o.ordertime BETWEEN s.shiptime - INTERVAL '4' HOUR AND s.shiptime
```

NOTE

1. Stream SQL Join supports only inner join.
2. The **ON** clause should include an equal join condition.
3. Time attributes support only the processing time and event time.
4. The window condition supports only the bounded time range, for example, **o.proctime BETWEEN s.proctime - INTERVAL '1' HOUR AND s.proctime + INTERVAL '1' HOUR**. The unbounded range such as **o. proctime > s.proctime** is not supported. The **proctime** attribute of two streams must be included. **o.proctime BETWEEN proctime () AND proctime () + 1** is not supported.

1.3.4.4.4 Flink CEP in SQL

Flink CEP in SQL

Flink allows users to represent complex event processing (CEP) query results in SQL for pattern matching and evaluate event streams on Flink engines.

SQL Query Syntax

CEP SQL is implemented through the **MATCH_RECOGNIZE** SQL syntax. The **MATCH_RECOGNIZE** clause is supported by Oracle SQL since Oracle Database 12c and is used to indicate event pattern matching in SQL. Apache Calcite also supports the **MATCH_RECOGNIZE** clause.

Flink uses Calcite to analyze SQL query results. Therefore, this operation complies with the Apache Calcite syntax.

```
MATCH_RECOGNIZE (
      [ PARTITION BY expression [, expression ]* ]
      [ ORDER BY orderItem [, orderItem ]* ]
      [ MEASURES measureColumn [, measureColumn ]* ]
      [ ONE ROW PER MATCH | ALL ROWS PER MATCH ]
      [ AFTER MATCH
            ( SKIP TO NEXT ROW
            | SKIP PAST LAST ROW
            | SKIP TO FIRST variable
            | SKIP TO LAST variable
            | SKIP TO variable )
      ]
      PATTERN ( pattern )
      [ WITHIN intervalLiteral ]
      [ SUBSET subsetItem [, subsetItem ]* ]
      DEFINE variable AS condition [, variable AS condition ]*
      )
```

The syntax elements of the **MATCH_RECOGNIZE** clause are defined as follows:

(Optional) **-PARTITION BY**: defines partition columns. This clause is optional. If this parameter is not defined, the parallelism 1 is used.

(Optional) **-ORDER BY**: defines the sequence of events in a data flow. The **ORDER BY** clause is optional. If it is ignored, non-deterministic sorting is used. Since the order of events is important in pattern matching, this clause should be specified in most cases.

(Optional) **-MEASURES**: specifies the attribute value of the successfully matched event.

(Optional) **-ONE ROW PER MATCH | ALL ROWS PER MATCH**: defines how to output the result. **ONE ROW PER MATCH** indicates that only one row is output for each matching. **ALL ROWS PER MATCH** indicates that one row is output for each matching event.

(Optional) **-AFTER MATCH**: specifies the start position for processing after the next pattern is successfully matched.

-PATTERN: defines the matching pattern as a regular expression. The following operators can be used in the **PATTERN** clause: join operators, quantifier operators (*, +, ?, {n}, {n,}, {n,m}, and {,m}), branch operators (vertical bar |), and differential operators ('{- -}').

(Optional) **-WITHIN**: outputs a pattern clause match only when the match occurs within the specified time.

(Optional) **-SUBSET**: combines one or more associated variables defined in the **DEFINE** clause.

-DEFINE: specifies the Boolean condition, which defines the variables used in the **PATTERN** clause.

In addition, the **MATCH_RECOGNIZE** clause supports the following functions:

-MATCH_NUMBER(): Used in the **MEASURES** clause to allocate the same number to each row that is successfully matched.

-CLASSIFIER(): Used in the **MEASURES** clause to indicate the mapping between matched rows and variables.

-FIRST() and **LAST()**: Used in the **MEASURES** clause to return the value of the expression evaluated in the first or last row of the row set mapped to the schema variable.

-NEXT() and **PREV()**: Used in the **DEFINE** clause to evaluate an expression using the previous or next row in a partition.

-RUNNING and **FINAL** keywords: Used to determine the semantics required for aggregation. **RUNNING** can be used in the **MEASURES** and **DEFINE** clauses, whereas **FINAL** can be used only in the **MEASURES** clause.

- Aggregate functions (**COUNT**, **SUM**, **AVG**, **MAX**, **MIN**): Used in the **MEASURES** and **DEFINE** clauses.

Query Example

The following query finds the V-shaped pattern in the stock price data flow.

```
SELECT *
  FROM MyTable
  MATCH_RECOGNIZE (
    ORDER BY rowtime
    MEASURES
      STRT.name as s_name,
      LAST(DOWN.name) as down_name,
      LAST(UP.name) as up_name
    ONE ROW PER MATCH
    PATTERN (STRT DOWN+ UP+)
    DEFINE
      DOWN AS DOWN.v < PREV(DOWN.v),
      UP AS UP.v > PREV(UP.v)
  )
```

In the following query, the aggregate function **AVG** is used in the **MEASURES** clause of **SUBSET E** consisting of variables related to A and C.

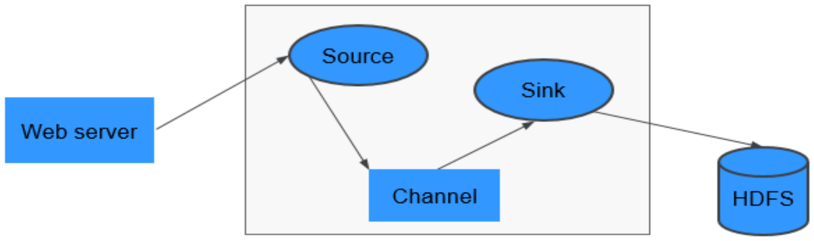
```
SELECT *
  FROM Ticker
  MATCH_RECOGNIZE (
    MEASURES
      AVG(E.price) AS avgPrice
    ONE ROW PER MATCH
    AFTER MATCH SKIP PAST LAST ROW
    PATTERN (A B+ C)
    SUBSET E = (A,C)
    DEFINE
      A AS A.price < 30,
      B AS B.price < 20,
      C AS C.price < 30
  )
```

1.3.5 Flume

1.3.5.1 Flume Basic Principles

Flume is a distributed, reliable, and HA system that supports massive log collection, aggregation, and transmission. Flume supports customization of various data senders in the log system for data collection. In addition, Flume can roughly process data and write data to various data receivers (customizable). A Flume-NG is a branch of Flume. It is simple, small, and easy to deploy. The following figure shows the basic architecture of the Flume-NG.

Figure 1-25 Flume-NG architecture



A Flume-NG consists of agents. Each agent consists of three components (source, channel, and sink). A source is used for receiving data. A channel is used for transmitting data. A sink is used for sending data to the next end.

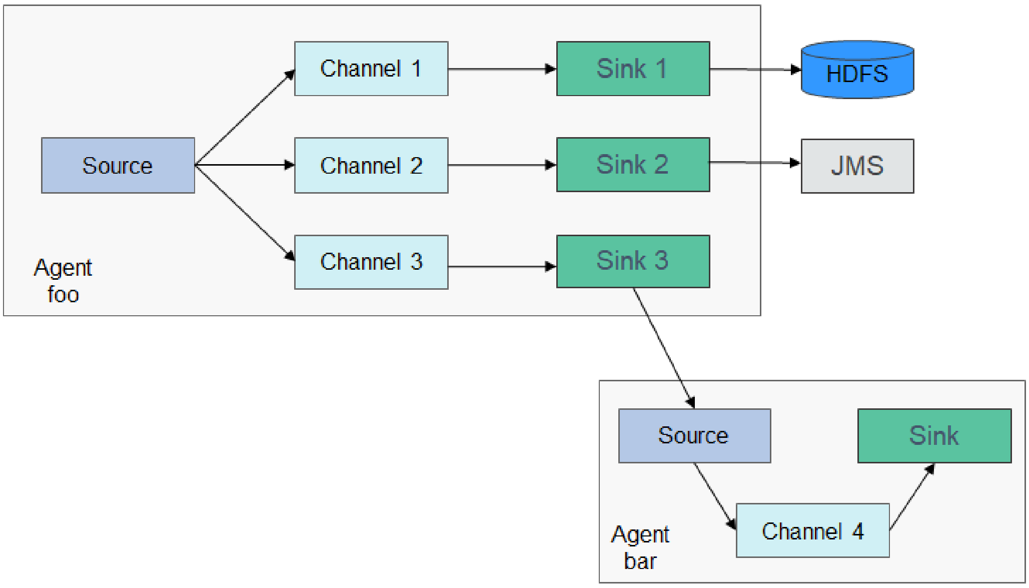
Table 1-4 Module description

Module	Description
Source	A source receives data or generates data by using a special mechanism, and places the data in batches in one or more channels. The source can work in data-driven or polling mode. Typical source types are as follows: ● Sources that are integrated with the system, such as Syslog and Netcat ● Sources that automatically generate events, such as Exec and SEQ ● IPC sources that are used for communication between agents, such as Avro A source must be associated with at least one channel.
Channel	A channel is used to buffer data between a source and a sink. The channel caches data from the source and deletes that data after the sink sends the data to the next-hop channel or final destination. Different channels provide different persistence levels. ● Memory channel: non-persistency ● File channel: Write-Ahead Logging (WAL)-based persistence ● JDBC channel: persistency implemented based on the embedded database The channel supports the transaction feature to ensure simple sequential operations. A channel can work with sources and sinks of any quantity.

Module	Description
Sink	A sink sends data to the next-hop channel or final destination. Once completed, the transmitted data is removed from the channel. Typical sink types are as follows: ● Sinks that send storage data to the final destination, such as HDFS and HBase ● Sinks that are consumed automatically, such as Null Sink ● IPC sinks used for communication between Agents, such as Avro A sink must be associated with a specific channel.

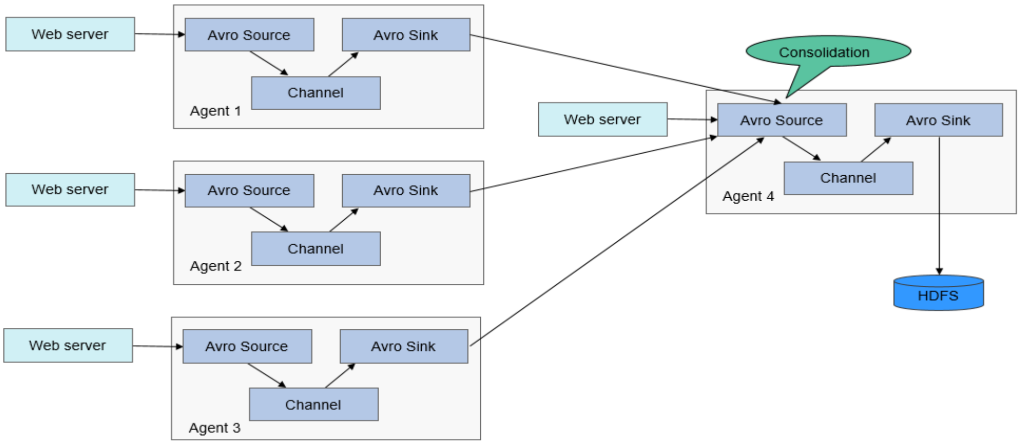
As shown in **Figure 1-26**, a Flume client can have multiple sources, channels, and sinks.

Figure 1-26 Flume structure



The reliability of Flume depends on transaction switchovers between agents. If the next agent breaks down, the channel stores data persistently and transmits data until the agent recovers. The availability of Flume depends on the built-in load balancing and failover mechanisms. Both the channel and agent can be configured with multiple entities between which they can use load balancing policies. Each agent is a Java Virtual Machine (JVM) process. A server can have multiple agents. Collection nodes (for example, Agents 1, 2, 3) process logs. Aggregation nodes (for example, Agent 4) write the logs into HDFS. The agent of each collection node can select multiple aggregation nodes for load balancing.

Figure 1-27 Flume cascading



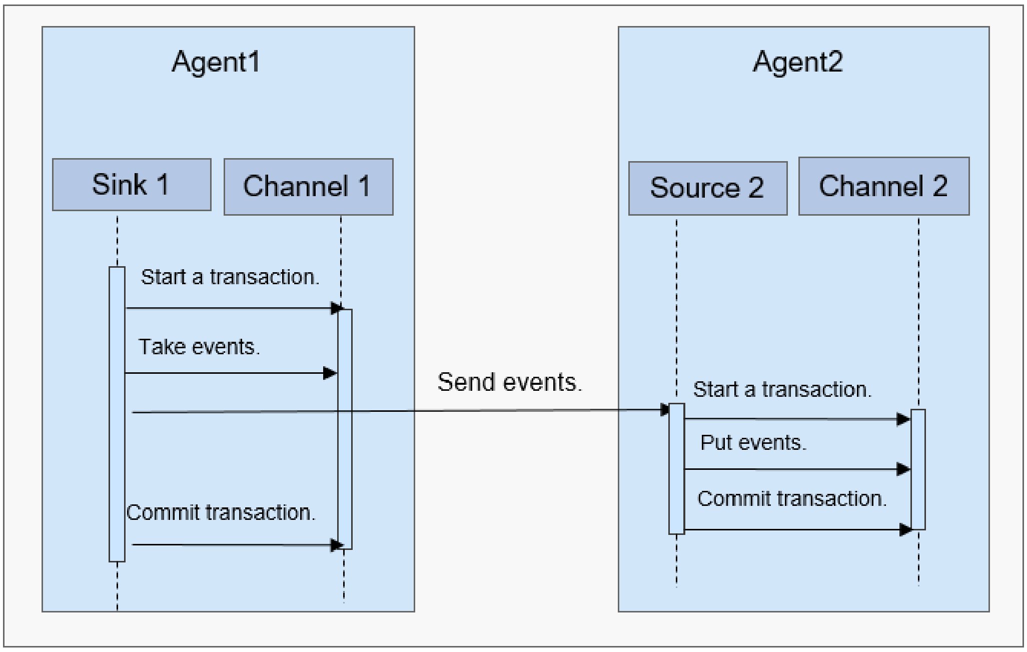
For details about Flume architecture and principles, see **https://flume.apache.org/releases/1.9.0.html**.

Principle

Reliability Between Agents

Figure 1-28 shows the data exchange between agents.

Figure 1-28 Data transmission process



1. Flume ensures reliable data transmission based on transactions. When data flows from one agent to another agent, the two transactions take effect. The sink of Agent 1 (agent that sends a message) needs to obtain a message from a channel and sends the message to Agent 2 (agent that receives the

message). If Agent 2 receives and successfully processes the message, Agent 1 will submit a transaction, indicating a successful and reliable data transmission.

2. When Agent 2 receives the message sent by Agent 1 and starts a new transaction, after the data is processed successfully (written to a channel), Agent 2 submits the transaction and sends a success response to Agent 1.
3. Before a commit operation, if the data transmission fails, the last transcription starts and retransmits the data that fails to be transmitted last time. The commit operation has written the transaction into a disk. Therefore, the last transaction can continue after the process fails and restores.

1.3.5.2 Relationship Between Flume and Other Components

Relationship Between Flume and HDFS

If HDFS is configured as the Flume sink, HDFS functions as the final data storage system of Flume. Flume installs, configures, and writes all transmitted data into HDFS.

Relationship Between Flume and HBase

If HBase is configured as the Flume sink, HBase functions as the final data storage system of Flume. Flume writes all transmitted data into HBase based on configurations.

1.3.5.3 Flume Enhanced Open Source Features

Flume Enhanced Open Source Features

- Improving transmission speed: Multiple lines instead of only one line of data can be specified as an event. This improves the efficiency of code execution and reduces the times of disk writes.
- Transferring ultra-large binary files: According to the current memory usage, Flume automatically adjusts the memory used for transferring ultra-large binary files to prevent out-of-memory.
- Supporting the customization of preparations before and after transmission: Flume supports customized scripts to be run before or after transmission for making preparations.
- Managing client alarms: Flume receives Flume client alarms through MonitorServer and reports the alarms to the alarm management center on MRS Manager.

1.3.6 HBase

1.3.6.1 HBase Basic Principles

HBase undertakes data storage. HBase is an open source, column-oriented, distributed storage system that is suitable for storing massive amounts of unstructured or semi-structured data. It features high reliability, high performance, and flexible scalability, and supports real-time data read/write. For more information about HBase, see **https://hbase.apache.org/**.

Typical features of a table stored in HBase are as follows:

- Big table (BigTable): One table contains hundred millions of rows and millions of columns.
- Column-oriented: Column-oriented storage, retrieval, and permission control
- Sparse: Null columns in the table do not occupy any storage space.

MRS HBase supports secondary indexing to allow indexes to be created for column values so that data can be filtered by column using native HBase APIs.

HBase Architecture

An HBase cluster consists of active and standby HMaster processes and multiple RegionServer processes.

Figure 1-29 HBase architecture

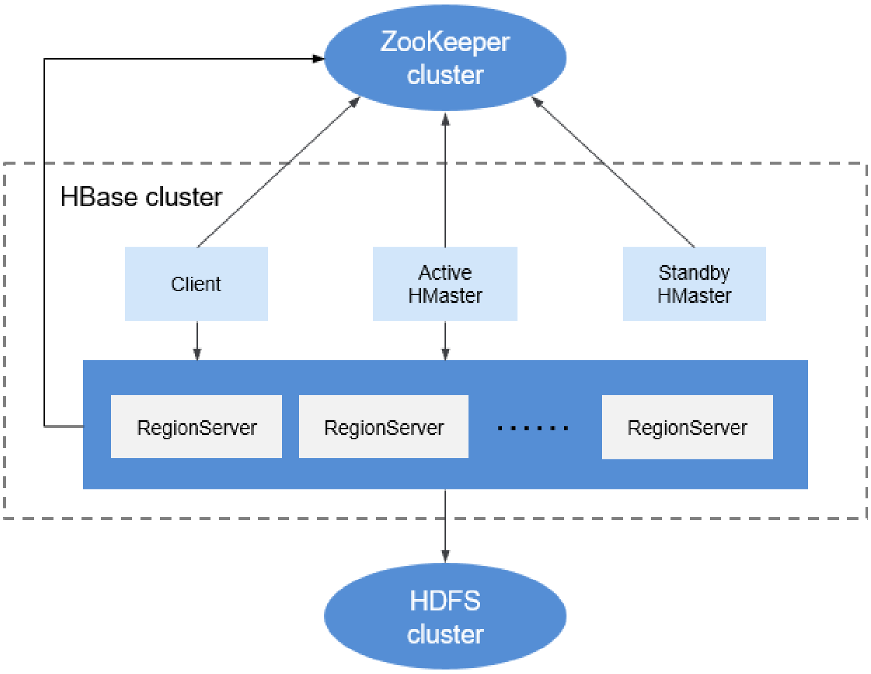


Table 1-5 Module description

Module	Description
Master	Master is also called HMaster. In HA mode, HMaster consists of an active HMaster and a standby HMaster. ● Active Master: manages RegionServer in HBase, including the creation, deletion, modification, and query of a table, balances the load of RegionServer, adjusts the distribution of Region, splits Region and distributes Region after it is split, and migrates Region after RegionServer expires. ● Standby Master: takes over services when the active HMaster is faulty. The original active HMaster demotes to the standby HMaster after the fault is rectified.
Client	Client communicates with Master for management and with RegionServer for data protection by using the Remote Procedure Call (RPC) mechanism of HBase.
RegionServer	RegionServer provides read and write services of table data as a data processing and computing unit in HBase. RegionServer is deployed with DataNodes of HDFS clusters to store data.
ZooKeeper cluster	ZooKeeper provides distributed coordination services for processes in HBase clusters. Each RegionServer is registered with ZooKeeper so that the active Master can obtain the health status of each RegionServer.
HDFS cluster	HDFS provides highly reliable file storage services for HBase. All HBase data is stored in the HDFS.

HBase Principles

- **HBase Data Model**

  HBase stores data in tables, as shown in **Figure 1-30**. Data in a table is divided into multiple Regions, which are allocated by Master to RegionServers for management.

  Each Region contains data within a RowKey range. An HBase data table contains only one Region at first. As the number of data increases and reaches the upper limit of the Region capacity, the Region is split into two Regions. You can define the RowKey range of a Region when creating a table or define the Region size in the configuration file.

Figure 1-30 HBase data model

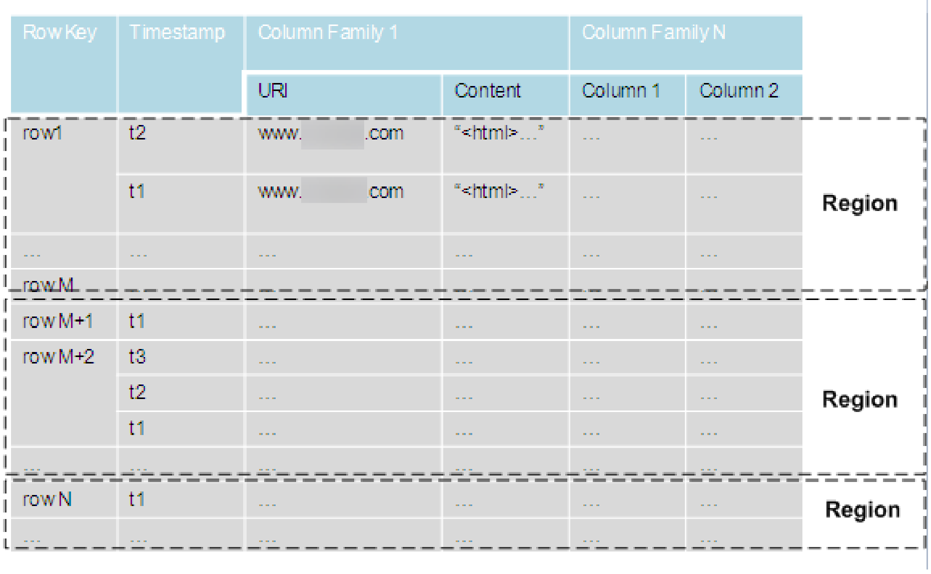


Table 1-6 Concepts

Module	Description
RowKey	Similar to the primary key in a relationship table, which is the unique ID of the data in each row. A RowKey can be a string, integer, or binary string. All records are stored after being sorted by RowKey.
Timestamp	The timestamp of a data operation. Data can be specified with different versions by time stamp. Data of different versions in each cell is stored by time in descending order.
Cell	Minimum storage unit of HBase, consisting of keys and values. A key consists of six fields, namely row, column family, column qualifier, timestamp, type, and MVCC version. Values are the binary data objects.
Column Family	One or multiple horizontal column families form a table. A column family can consist of multiple random columns. A column is a label under a column family, which can be added as required when data is written. The column family supports dynamic expansion so the number and type of columns do not need to be predefined. Columns of a table in HBase are sparsely distributed. The number and type of columns in different rows can be different. Each column family has the independent time to live (TTL). You can lock the row only. Operations on the row in a column family are the same as those on other rows.
Column	Similar to traditional databases, HBase tables also use columns to store data of the same type.

- **RegionServer Data Storage**

RegionServer manages the regions allocated by HMaster. **Figure 1-31** shows the data storage structure of RegionServer.

Figure 1-31 RegionServer data storage structure

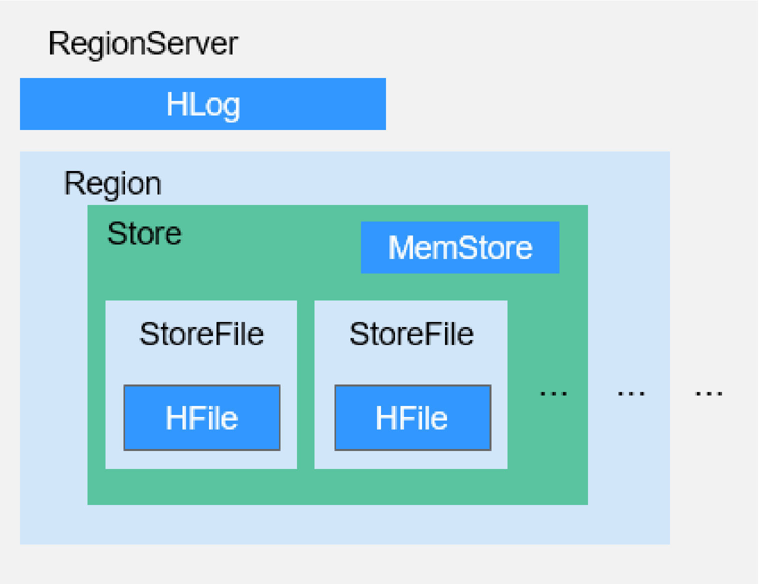


Table 1-7 lists each component of Region described in **Figure 1-31**.

Table 1-7 Region structure description

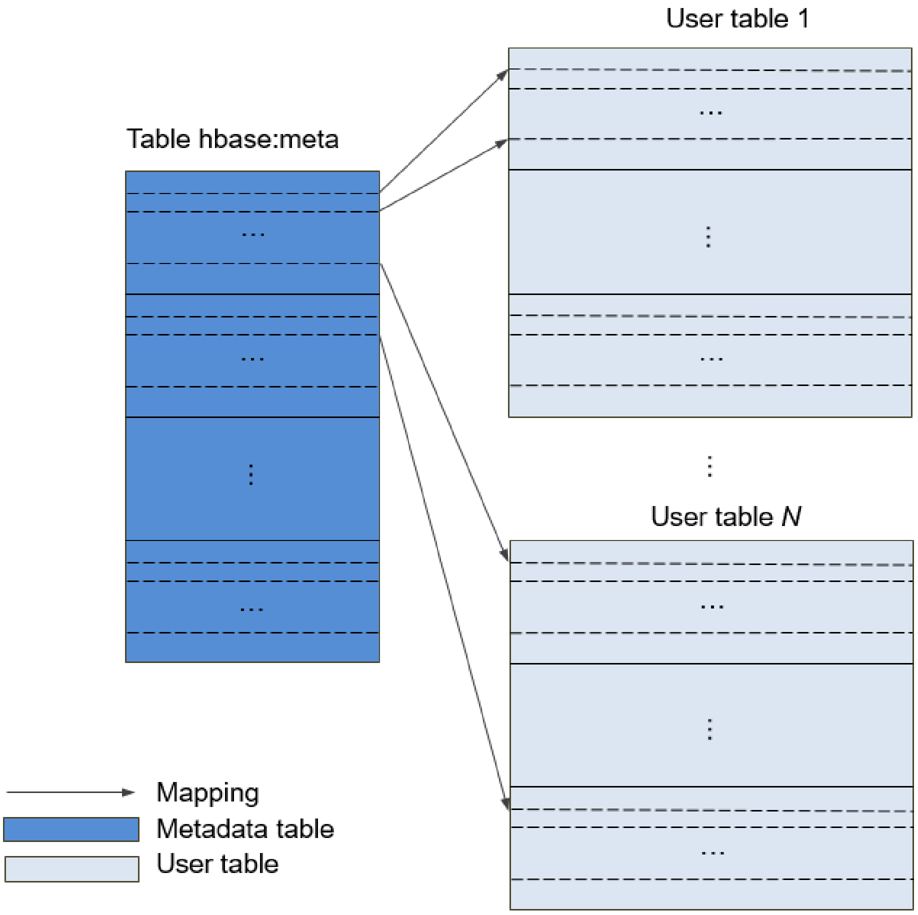
Module	Description
Store	A Region consists of one or multiple Stores. Each Store maps a column family in **Figure 1-30**.
MemStore	A Store contains one MemStore. The MemStore caches data inserted to a Region by the client. When the MemStore capacity reaches the upper limit, RegionServer flushes data in MemStore to the HDFS.
StoreFile	The data flushed to the HDFS is stored as a StoreFile in the HDFS. As more data is inserted, multiple StoreFiles are generated in a Store. When the number of StoreFiles reaches the upper limit, RegionServer merges multiple StoreFiles into a big StoreFile.
HFile	HFile defines the storage format of StoreFiles in a file system. HFile is the underlying implementation of StoreFile.
HLog	HLogs prevent data loss when RegionServer is faulty. Multiple Regions in a RegionServer share the same HLog.

- **Metadata Table**

  The metadata table is a special HBase table, which is used by the client to locate a region. Metadata table includes **hbase:meta** table to record region information of user tables, such as the region location and start and end RowKey.

Figure 1-32 shows the mapping relationship between metadata tables and user tables.

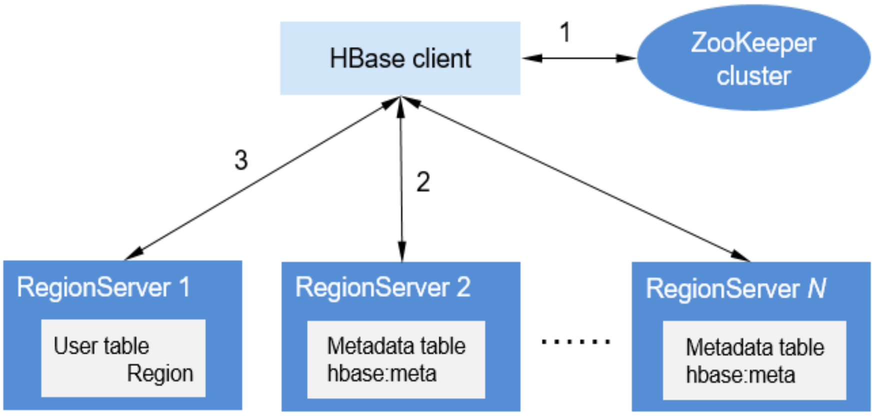
Figure 1-32 Mapping relationships between metadata tables and user tables



- **Data Operation Process**

  Figure 1-33 shows the HBase data operation process.

Figure 1-33 Data processing



a. When you add, delete, modify, and query HBase data, the HBase client first connects to ZooKeeper to obtain information about the RegionServer where the **hbase:meta** table is located. If you modify the namespace, such as creating and deleting a table, you need to access HMaster to update the meta information.
b. The HBase client connects to the RegionServer where the region of the **hbase:meta** table is located and obtains the RegionServer location where the region of the user table resides.
c. Then the HBase client connects to the RegionServer where the region of the user table is located and issues a data operation command to the RegionServer. The RegionServer executes the command.

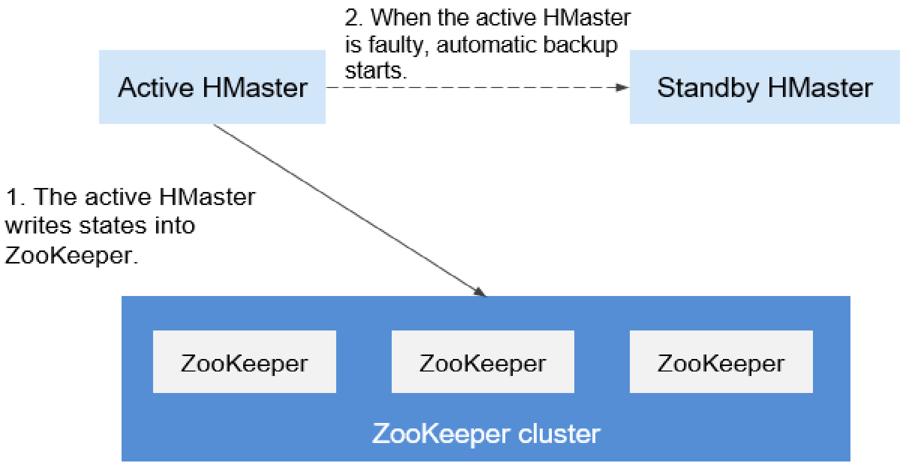
To improve data processing efficiency, the HBase client caches region information of the **hbase:meta** table and user table. When an application initiates a second data operation, the HBase client queries the region information from the memory. If no match is found in the memory, the HBase client performs the preceding operations to obtain region information.

1.3.6.2 HBase HA Solution

HBase HA

HMaster in HBase allocates Regions. When one RegionServer service is stopped, HMaster migrates the corresponding Region to another RegionServer. The HMaster HA feature is brought in to prevent HBase functions from being affected by the HMaster single point of failure (SPOF).

Figure 1-34 HMaster HA implementation architecture


2. When the active HMaster is faulty, automatic backup starts.
Active HMaster
Standby HMaster
1. The active HMaster writes states into ZooKeeper.
ZooKeeper
ZooKeeper
ZooKeeper
ZooKeeper cluster

The HMaster HA architecture is implemented by creating the ephemeral ZooKeeper node in a ZooKeeper cluster.

Upon startup, HMaster nodes try to create a master znode in the ZooKeeper cluster. The HMaster node that creates the master znode first becomes the active HMaster, and the other is the standby HMaster.

It will add watch events to the master node. If the service on the active HMaster is stopped, the active HMaster disconnects from the ZooKeeper cluster. After the session expires, the active HMaster disappears. The standby HMaster detects the disappearance of the active HMaster through watch events and creates a master node to make itself be the active one. Then, the active/standby switchover completes. If the failed node detects existence of the master node after being restarted, it enters the standby state and adds watch events to the master node.

When the client accesses the HBase, it first obtains the HMaster's address based on the master node information on the ZooKeeper and then establishes a connection to the active HMaster.

1.3.6.3 Relationship with Other Components

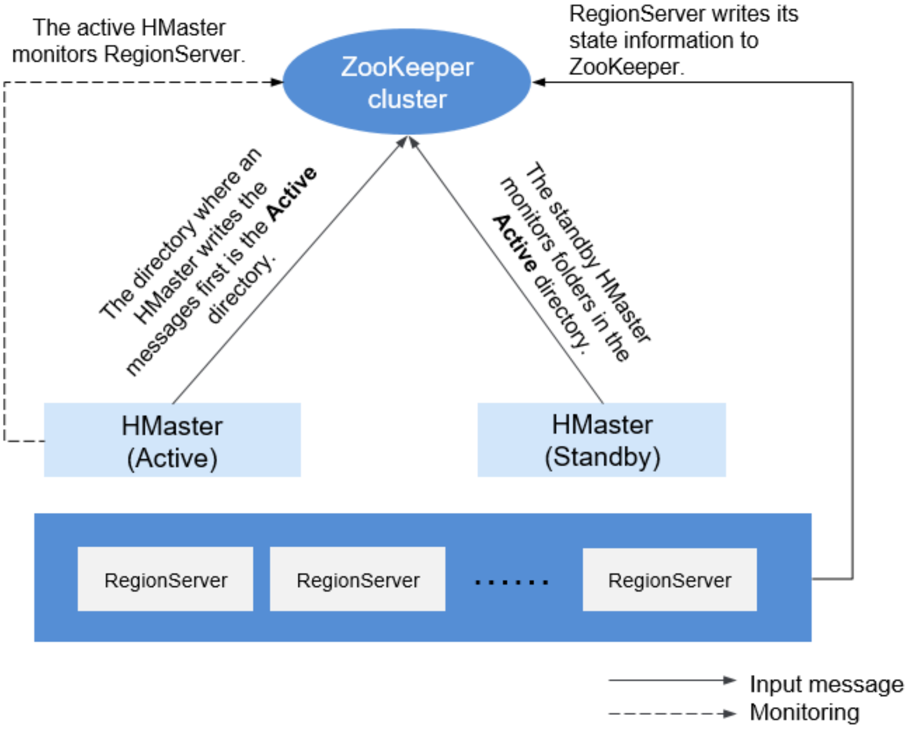
Relationship Between HDFS and HBase

HDFS is the subproject of Apache Hadoop. HBase uses the Hadoop Distributed File System (HDFS) as the file storage system. HBase is located in structured storage layer. The HDFS provides highly reliable support for lower-layer storage of HBase. All the data files of HBase can be stored in the HDFS, except some log files generated by HBase.

Relationship Between ZooKeeper and HBase

Figure 1-35 describes the relationship between ZooKeeper and HBase.

Figure 1-35 Relationship between ZooKeeper and HBase


The active HMaster monitors RegionServer.
ZooKeeper cluster
RegionServer writes its state information to ZooKeeper.
The directory where an HMaster writes the messages first is the **Active** directory.
The standby HMaster monitors folders in the **Active** directory.
HMaster (Active)
HMaster (Standby)
RegionServer
RegionServer
......
RegionServer
Input message
Monitoring

1. HRegionServer registers itself to ZooKeeper in Ephemeral node. ZooKeeper stores the HBase information, including the HBase metadata and HMaster addresses.
2. HMaster detects the health status of each HRegionServer using ZooKeeper, and monitors them.

3. HBase can deploy multiple HMasters (like HDFS NameNode). When the active HMatser node is faulty, the standby HMaster node obtains the state information of the entire cluster using ZooKeeper, which means that HBase single point faults can be avoided using ZooKeeper.

1.3.6.4 HBase Enhanced Open Source Features

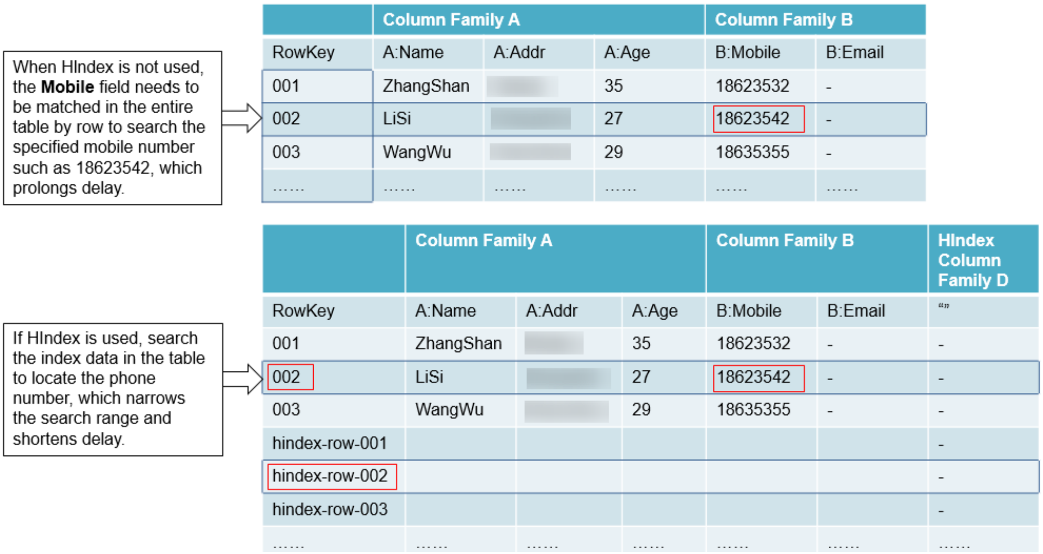
HIndex

HBase is a distributed storage database of the Key-Value type. Data of a table is sorted in the alphabetic order based on row keys. If you query data based on a specified row key or scan data in the scale of a specified row key, HBase can quickly locate the target data, enhancing the efficiency.

However, in most actual scenarios, you need to query the data of which the column value is *XXX*. HBase provides the Filter feature to query data with a specific column value. All data is scanned in the order of row keys, and then the data is matched with the specific column value until the required data is found. The Filter feature scans some unnecessary data to obtain the only required data. Therefore, the Filter feature cannot meet the requirements of frequent queries with high performance standards.

HBase HIndex is designed to address these issues. HBase HIndex enables HBase to query data based on specific column values.

Figure 1-36 HIndex



When HIndex is not used, the **Mobile** field needs to be matched in the entire table by row to search the specified mobile number such as 18623542, which prolongs delay.

	Column Family A			Column Family B	
RowKey	A:Name	A:Addr	A:Age	B:Mobile	B:Email
001	ZhangShan	[redacted]	35	18623532	-
002	LiSi	[redacted]	27	18623542	-
003	WangWu	[redacted]	29	18635355	-
……	……	……	……	……	……

If HIndex is used, search the index data in the table to locate the phone number, which narrows the search range and shortens delay.

	Column Family A			Column Family B		HIndex Column Family D
RowKey	A:Name	A:Addr	A:Age	B:Mobile	B:Email	“”
001	ZhangShan	[redacted]	35	18623532	-	-
002	LiSi	[redacted]	27	18623542	-	-
003	WangWu	[redacted]	29	18635355	-	-
hindex-row-001						-
hindex-row-002						-
hindex-row-003						-
……	……	……	……	……	……	……

- Rolling upgrade is not supported for index data.
- Restrictions of combined indexes:
  - All columns involved in combined indexes must be entered or deleted in a single mutation. Otherwise, inconsistency will occur.

    Index: **IDX1=>cf1:[q1->datatype],[q2];cf2:[q2->datatype]**

    Correct write operations:

    ```
    Put put = new Put(Bytes.toBytes("row"));
    put.addColumn(Bytes.toBytes("cf1"), Bytes.toBytes("q1"), Bytes.toBytes("valueA"));
    ```

```
put.addColumn(Bytes.toBytes("cf1"), Bytes.toBytes("q2"), Bytes.toBytes("valueB"));
put.addColumn(Bytes.toBytes("cf2"), Bytes.toBytes("q2"), Bytes.toBytes("valueC"));
table.put(put);
```

Incorrect write operations:

```
Put put1 = new Put(Bytes.toBytes("row"));
put1.addColumn(Bytes.toBytes("cf1"), Bytes.toBytes("q1"), Bytes.toBytes("valueA"));
table.put(put1);
Put put2 = new Put(Bytes.toBytes("row"));
put2.addColumn(Bytes.toBytes("cf1"), Bytes.toBytes("q2"), Bytes.toBytes("valueB"));
table.put(put2);
Put put3 = new Put(Bytes.toBytes("row"));
put3.addColumn(Bytes.toBytes("cf2"), Bytes.toBytes("q2"), Bytes.toBytes("valueC"));
table.put(put3);
```

  - The combined conditions-based query is supported only when the combined index column contains filter criteria, or StartRow and StopRow are not specified for some index columns.

    Index: **IDX1=>cf1:[q1->datatype],[q2];cf2:[q1->datatype]**

    Correct query operations:

```
scan 'table', {FILTER=>"SingleColumnValueFilter('cf1','q1',>=,'binary:valueA',true,true) AND
SingleColumnValueFilter('cf1','q2',>=,'binary:valueB',true,true) AND
SingleColumnValueFilter('cf2','q1',>=,'binary:valueC',true,true) "}

scan 'table', {FILTER=>"SingleColumnValueFilter('cf1','q1',=,'binary:valueA',true,true) AND
SingleColumnValueFilter('cf1','q2',>=,'binary:valueB',true,true)" }

scan 'table', {FILTER=>"SingleColumnValueFilter('cf1','q1',>=,'binary:valueA',true,true) AND
SingleColumnValueFilter('cf1','q2',>=,'binary:valueB',true,true) AND
SingleColumnValueFilter('cf2','q1',>=,'binary:valueC',true,true)",STARTROW=>'row001',STOPROW
=>'row100'}
```

    Incorrect query operations:

```
scan 'table', {FILTER=>"SingleColumnValueFilter('cf1','q1',>=,'binary:valueA',true,true) AND
SingleColumnValueFilter('cf1','q2',>=,'binary:valueB',true,true) AND
SingleColumnValueFilter('cf2','q1',>=,'binary:valueC',true,true)  AND
SingleColumnValueFilter('cf2','q2',>=,'binary:valueD',true,true)"}

scan 'table', {FILTER=>"SingleColumnValueFilter('cf1','q1',=,'binary:valueA',true,true) AND
SingleColumnValueFilter('cf2','q1',>=,'binary:valueC',true,true)" }

scan 'table', {FILTER=>"SingleColumnValueFilter('cf1','q1',=,'binary:valueA',true,true) AND
SingleColumnValueFilter('cf2','q2',>=,'binary:valueD',true,true)" }

scan 'table', {FILTER=>"SingleColumnValueFilter('cf1','q1',=,'binary:valueA',true,true) AND
SingleColumnValueFilter('cf1','q2',>=,'binary:valueB',true,true)" ,STARTROW=>'row001',STOPROW
=>'row100' }
```

- Do not explicitly configure any split policy for tables with index data.
- Other mutation operations, such as **increment** and **append**, are not supported.
- Index of the column with **maxVersions** greater than 1 is not supported.
- The data index column in a row cannot be updated.

  Index 1: **IDX1=>cf1:[q1->datatype],[q2];cf2:[q1->datatype]**

  Index 2: **IDX2=>cf2:[q2->datatype]**

  Correct update operations:

```
Put put1 = new Put(Bytes.toBytes("row"));
put1.addColumn(Bytes.toBytes("cf1"), Bytes.toBytes("q1"), Bytes.toBytes("valueA"));
put1.addColumn(Bytes.toBytes("cf1"), Bytes.toBytes("q2"), Bytes.toBytes("valueB"));
put1.addColumn(Bytes.toBytes("cf2"), Bytes.toBytes("q1"), Bytes.toBytes("valueC"));
put1.addColumn(Bytes.toBytes("cf2"), Bytes.toBytes("q2"), Bytes.toBytes("valueD"));
table.put(put1);
```

```
Put put2 = new Put(Bytes.toBytes("row"));
put2.addColumn(Bytes.toBytes("cf1"), Bytes.toBytes("q3"), Bytes.toBytes("valueE"));
put2.addColumn(Bytes.toBytes("cf2"), Bytes.toBytes("q3"), Bytes.toBytes("valueF"));
table.put(put2);
```

Incorrect update operations:

```
Put put1 = new Put(Bytes.toBytes("row"));
put1.addColumn(Bytes.toBytes("cf1"), Bytes.toBytes("q1"), Bytes.toBytes("valueA"));
put1.addColumn(Bytes.toBytes("cf1"), Bytes.toBytes("q2"), Bytes.toBytes("valueB"));
put1.addColumn(Bytes.toBytes("cf2"), Bytes.toBytes("q1"), Bytes.toBytes("valueC"));
put1.addColumn(Bytes.toBytes("cf2"), Bytes.toBytes("q2"), Bytes.toBytes("valueD"));
table.put(put1);

Put put2 = new Put(Bytes.toBytes("row"));
put2.addColumn(Bytes.toBytes("cf1"), Bytes.toBytes("q1"), Bytes.toBytes("valueA_new"));
put2.addColumn(Bytes.toBytes("cf1"), Bytes.toBytes("q2"), Bytes.toBytes("valueB_new"));
put2.addColumn(Bytes.toBytes("cf2"), Bytes.toBytes("q1"), Bytes.toBytes("valueC_new"));
put2.addColumn(Bytes.toBytes("cf2"), Bytes.toBytes("q2"), Bytes.toBytes("valueD_new"));
table.put(put2);
```

- The table to which an index is added cannot contain a value greater than 32 KB.
- If user data is deleted due to the expiration of the column-level TTL, the corresponding index data is not deleted immediately. It will be deleted in the major compaction operation.
- The TTL of the user column family cannot be modified after the index is created.
  - If the TTL of a column family increases after an index is created, delete the index and re-create one. Otherwise, some generated index data will be deleted before user data is deleted.
  - If the TTL value of the column family decreases after an index is created, the index data will be deleted after user data is deleted.
- The index query does not support the reverse operation, and the query results are disordered.
- The index does not support the **clone snapshot** operation.
- The index table must use HIndexWALPlayer to replay logs. WALPlayer cannot be used to replay logs.

```
hbase org.apache.hadoop.hbase.hindex.mapreduce.HIndexWALPlayer
Usage: WALPlayer [options] <wal inputdir> <tables> [<tableMappings>]
Read all WAL entries for <tables>.
If no tables ("") are specific, all tables are imported.
(Careful, even -ROOT- and hbase:meta entries will be imported in that case.)
Otherwise <tables> is a comma separated list of tables.

The WAL entries can be mapped to new set of tables via <tableMapping>.
<tableMapping> is a command separated list of targettables.
If specified, each table in <tables> must have a mapping.

By default WALPlayer will load data directly into HBase.
To generate HFiles for a bulk data load instead, pass the option:
  -Dwal.bulk.output=/path/for/output
  (Only one table can be specified, and no mapping is allowed!)
Other options: (specify time range to WAL edit to consider)
  -Dwal.start.time=[date|ms]
  -Dwal.end.time=[date|ms]
For performance also consider the following options:
  -Dmapreduce.map.speculative=false
  -Dmapreduce.reduce.speculative=false
```

- When the **deleteall** command is executed for the index table, the performance is low.

- The index table does not support HBCK. To use HBCK to repair the index table, delete the index data first.

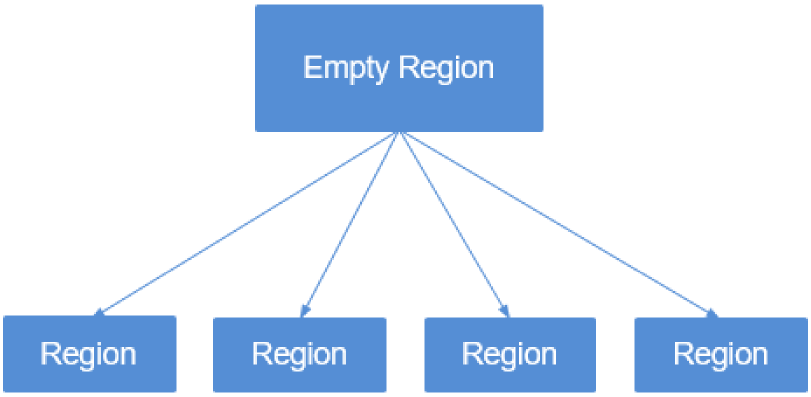
Multi-point Division

When you create tables that are pre-divided by region in HBase, you may not know the data distribution trend so the division by region may be inappropriate. After the system runs for a period, regions need to be divided again to achieve better performance. Only empty regions can be divided.

The region division function delivered with HBase divides regions only when they reach the threshold. This is called "single point division".

To achieve better performance when regions are divided based on user requirements, multi-point division is developed, which is also called "dynamic division". That is, an empty region is pre-divided into multiple regions to prevent performance deterioration caused by insufficient region space.

Figure 1-37 Multi-point division



Connection Limitation

Too many sessions mean that too many queries and MapReduce tasks are running on HBase, which compromises HBase performance and even causes service rejection. You can configure parameters to limit the maximum number of sessions that can be established between the client and the HBase server to achieve HBase overload protection.

Improved Disaster Recovery

The disaster recovery (DR) capabilities between the active and standby clusters can enhance HA of the HBase data. The active cluster provides data services and the standby cluster backs up data. If the active cluster is faulty, the standby cluster takes over data services. Compared with the open source replication function, this function is enhanced as follows:

1. The standby cluster whitelist function is only applicable to pushing data to a specified cluster IP address.
2. In the open source version, replication is synchronized based on WAL, and data backup is implemented by replaying WAL in the standby cluster. For

BulkLoad operations, since no WAL is generated, data will not be replicated to the standby cluster. By recording BulkLoad operations on the WAL and synchronizing them to the standby cluster, the standby cluster can read BulkLoad operation records through WAL and load HFile in the active cluster to the standby cluster to implement data backup.

3. In the open source version, HBase filters ACLs. Therefore, ACL information will not be synchronized to the standby cluster. By adding a filter (**org.apache.hadoop.hbase.replication.SystemTableWALEntryFilterAllowACL**), ACL information can be synchronized to the standby cluster. You can configure **hbase.replication.filter.sytemWALEntryFilter** to enable the filter and implement ACL synchronization.
4. As for read-only restriction of the standby cluster, only super users within the standby cluster can modify the HBase of the standby cluster. In other words, HBase clients outside the standby cluster can only read the HBase of the standby cluster.

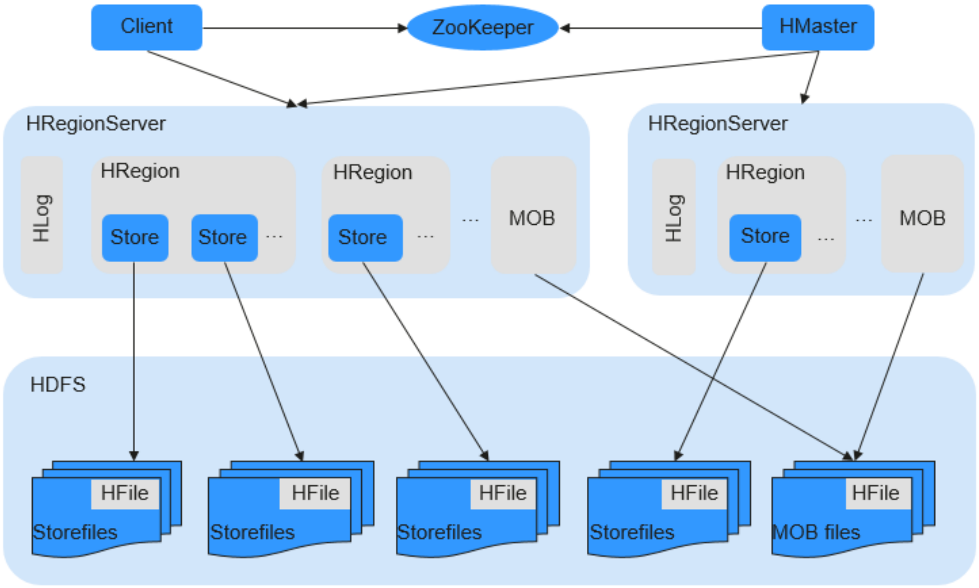
HBase MOB

In the actual application scenarios, data in various sizes needs to be stored, for example, image data and documents. Data whose size is smaller than 10 MB can be stored in HBase. HBase can yield the best read-and-write performance for data whose size is smaller than 100 KB. If the size of data stored in HBase is greater than 100 KB or even reaches 10 MB and the same number of data files are inserted, the total data amount is large, causing frequent compaction and split, high CPU consumption, high disk I/O frequency, and low performance.

MOB data (whose size ranges from 100 KB to 10 MB) is stored in a file system (for example, HDFS) in HFile format. The expiredMobFileCleaner and Sweeper tools are used to manage HFiles and save the address and size information about the HFiles to the store of HBase as values. This greatly decreases the compaction and split frequency in HBase and improves performance.

As shown in **Figure 1-38**, MOB indicates mobstore stored on HRegion. Mobstore stores keys and values. Wherein, a key is the corresponding key in HBase, and a value is the reference address and data offset stored in the file system. When reading data, mobstore uses its own scanner to read key-value data objects and uses the address and data size information in the value to obtain target data from the file system.

Figure 1-38 MOB data storage principle



HFS

HBase FileStream (HFS) is an independent HBase file storage module. It is used in MRS upper-layer applications by encapsulating HBase and HDFS interfaces to provide these upper-layer applications with functions such as file storage, read, and deletion.

In the Hadoop ecosystem, the HDFS and HBase face tough problems in mass file storage in some scenarios:

- If a large number of small files are stored in HDFS, the NameNode will be under great pressure.
- Some large files cannot be directly stored on HBase due to HBase APIs and internal mechanisms.

HFS is developed for the mixed storage of massive small files and some large files in Hadoop. Simply speaking, massive small files (smaller than 10 MB) and some large files (greater than 10 MB) need to be stored in HBase tables.

For such a scenario, HFS provides unified operation APIs similar to HBase function APIs.

Multiple RegionServers Deployed on the Same Server

Multiple RegionServers can be deployed on one node to improve HBase resource utilization.

If only one RegionServer is deployed, resource utilization is low due to the following reasons:

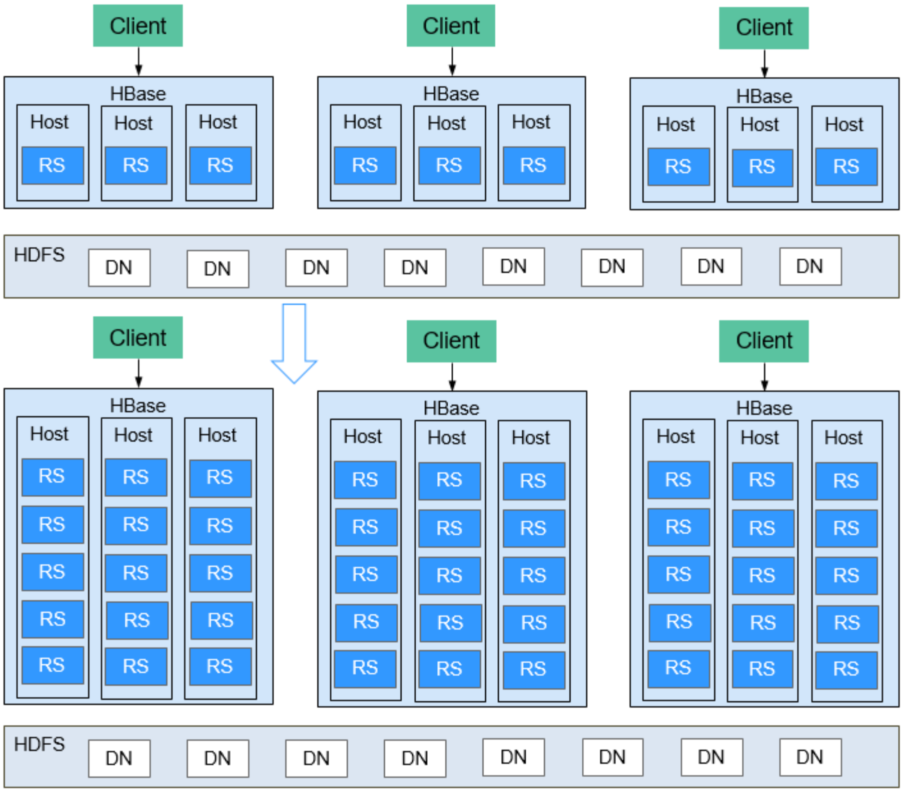
1. A RegionServer supports a limited number of regions, and therefore memory and CPU resources cannot be fully used.
2. A single RegionServer supports a maximum of 20 TB data, of which two copies require 40 TB, and three copies require 60 TB. In this case, 96 TB capacity cannot be used up.

3. Poor write performance: One RegionServer is deployed on a physical server, and only one HLog exists. Only three disks can be written at the same time.

The HBase resource utilization can be improved when multiple RegionServers are deployed on the same server.

1. A physical server can be configured with a maximum of five RegionServers. The number of RegionServers deployed on each physical server can be configured as required.
2. Resources such as memory, disks, and CPUs can be fully used.
3. A physical server supports a maximum of five HLogs and allows data to be written to 15 disks at the same time, significantly improving write performance.

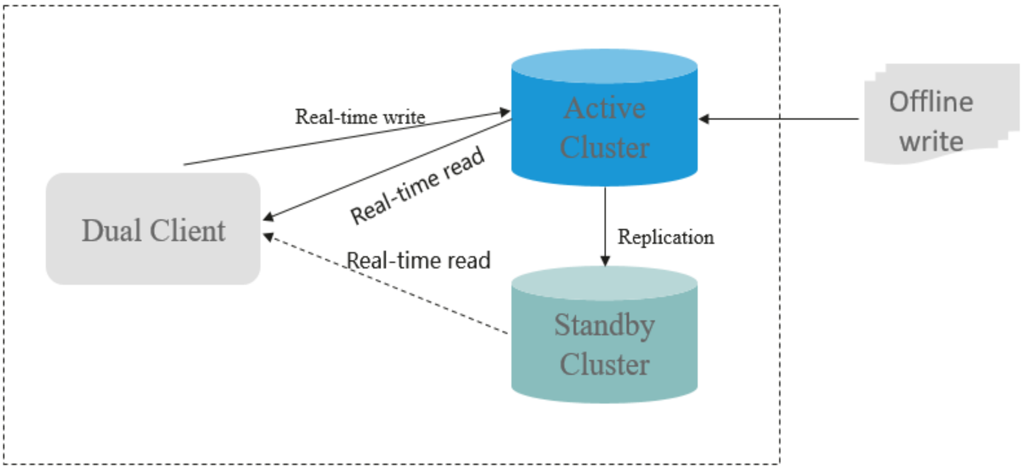
Figure 1-39 Improved HBase resource utilization



HBase Dual-Read

In the HBase storage scenario, it is difficult to ensure 99.9% query stability due to GC, network jitter, and bad sectors of disks. The HBase dual-read feature is added to meet the requirements of low glitches during large-data-volume random read.

The HBase dual-read feature is based on the DR capability of the active and standby clusters. The probability that the two clusters generate glitches at the same time is far less than that of one cluster. The dual-cluster concurrent access mode is used to ensure query stability. When a user initiates a query request, the HBase service of the two clusters is queried at the same time. If the active cluster does not return any result after a period of time (the maximum tolerable glitch time), the data of the cluster with the fastest response can be used. The following figure shows the working principle.



1.3.7 HDFS

1.3.7.1 HDFS Basic Principles

Hadoop Distributed File System (HDFS) implements reliable and distributed read/write of massive amounts of data. HDFS is applicable to the scenario where data read/write features "write once and read multiple times". However, the write operation is performed in sequence, that is, it is a write operation performed during file creation or an adding operation performed behind the existing file. HDFS ensures that only one caller can perform write operation on a file but multiple callers can perform read operation on the file at the same time.

Architecture

HDFS consists of active and standby NameNodes and multiple DataNodes, as shown in **Figure 1-40**.

HDFS works in master/slave architecture. NameNodes run on the master (active) node, and DataNodes run on the slave (standby) node. ZKFC should run along with the NameNodes.

The communication between NameNodes and DataNodes is based on Transmission Control Protocol (TCP)/Internet Protocol (IP). The NameNode, DataNode, ZKFC, and JournalNode can be deployed on Linux servers.

Figure 1-40 HA HDFS architecture

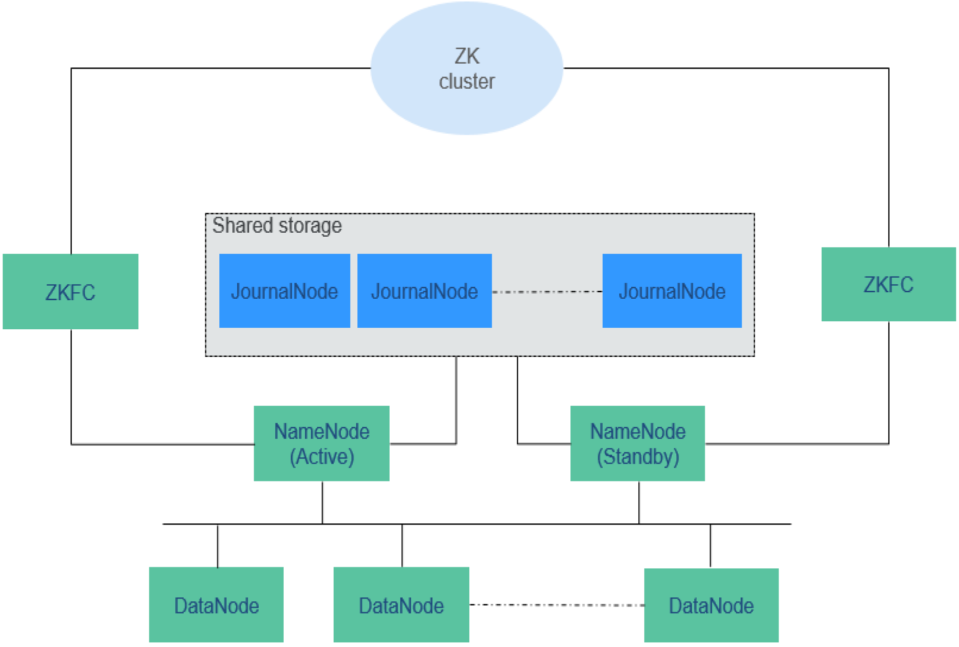


Table 1-8 describes the functions of each module shown in **Figure 1-40**.

Table 1-8 Module description

Module	Description
NameNode	A NameNode is used to manage the namespace, directory structure, and metadata information of a file system and provide the backup mechanism. The NameNode is classified into the following two types: ● Active NameNode: manages the namespace, maintains the directory structure and metadata of file systems, and records the mapping relationships between data blocks and files to which the data blocks belong. ● Standby NameNode: synchronizes with the data in the active NameNode, and takes over services from the active NameNode when the active NameNode is faulty. ● Observer NameNode: synchronizes with the data in the active NameNode, and processes read requests from the client.
DataNode	A DataNode is used to store data blocks of each file and periodically report the storage status to the NameNode.
JournalNode	In HA cluster, synchronizes metadata between the active and standby NameNodes.
ZKFC	ZKFC must be deployed for each NameNode. It monitors NameNode status and writes status information to ZooKeeper. ZKFC also has permissions to select the active NameNode.

Module	Description
ZK Cluster	ZooKeeper is a coordination service which helps the ZKFC to elect the active NameNode.
HttpFS gateway	HttpFS is a single stateless gateway process which provides the WebHDFS REST API for external processes and FileSystem API for the HDFS. HttpFS is used for data transmission between different versions of Hadoop. It is also used as a gateway to access the HDFS behind a firewall.

- **HDFS HA Architecture**

  HA is used to resolve the SPOF problem of NameNode. This feature provides a standby NameNode for the active NameNode. When the active NameNode is faulty, the standby NameNode can quickly take over to continuously provide services for external systems.

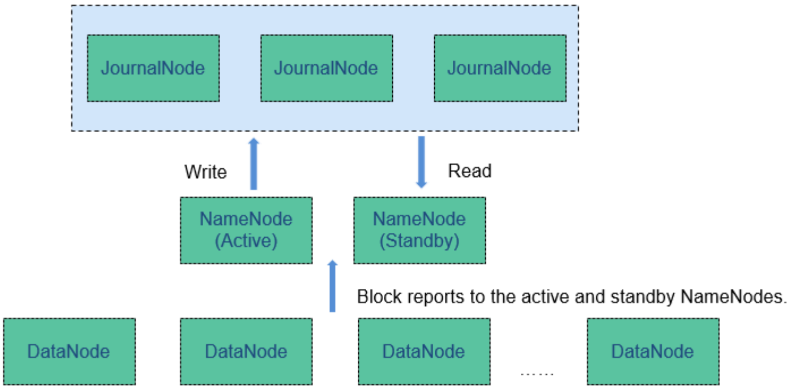
  In a typical HDFS HA scenario, there are usually two NameNodes. One is in the active state, and the other in the standby state.

  A shared storage system is required to support metadata synchronization of the active and standby NameNodes. This version provides Quorum Journal Manager (QJM) HA solution, as shown in **Figure 1-41**. A group of JournalNodes are used to synchronize metadata between the active and standby NameNodes.

  Generally, an odd number (2N+1) of JournalNodes are configured, and at least three JournalNodes are required. For one metadata update message, data writing is considered successful as long as data writing is successful on N +1 JournalNodes. In this case, data writing failure of a maximum of N JournalNodes is allowed. For example, when there are three JournalNodes, data writing failure of one JournalNode is allowed; when there are five JournalNodes, data writing failure of two JournalNodes is allowed.

  JournalNode is a lightweight daemon process and shares a host with other services of Hadoop. It is recommended that the JournalNode be deployed on the control node to prevent data writing failure on the JournalNode during massive data transmission.

  Figure 1-41 QJM-based HDFS architecture

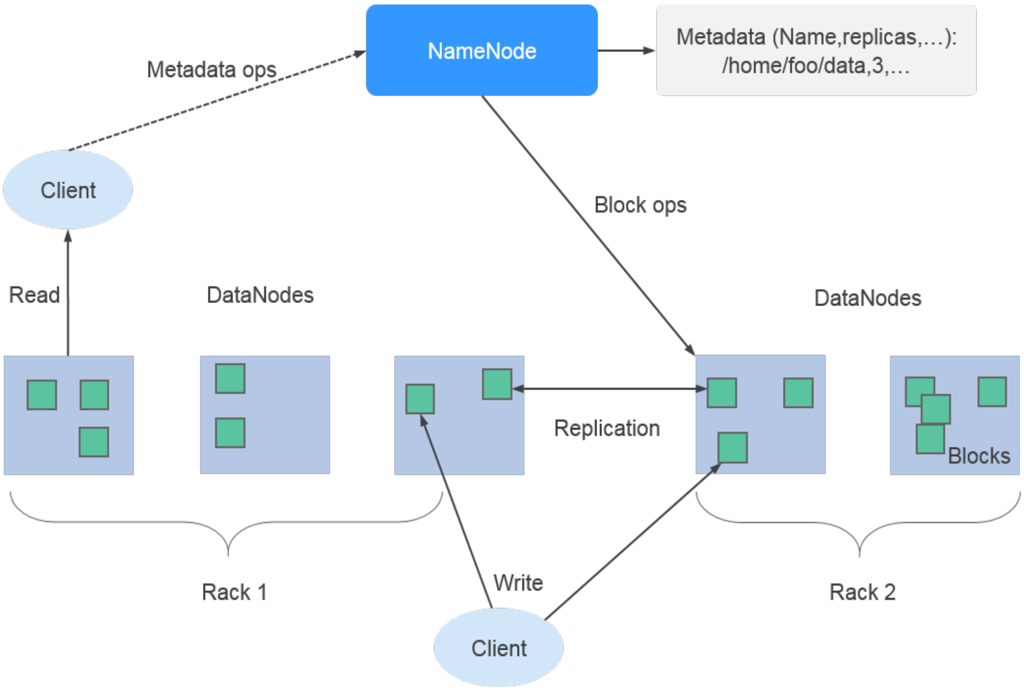
  

Principle

MRS uses the HDFS copy mechanism to ensure data reliability. One backup file is automatically generated for each file saved in HDFS, that is, two copies are generated in total. The number of HDFS copies can be queried using the **dfs.replication** parameter.

- When the Core node specification of the MRS cluster is set to non-local hard disk drive (HDD) and the cluster has only one Core node, the default number of HDFS copies is 1. If the number of Core nodes in the cluster is greater than or equal to 2, the default number of HDFS copies is 2.
- When the Core node specification of the MRS cluster is set to local disk and the cluster has only one Core node, the default number of HDFS copies is 1. If there are two Core nodes in the cluster, the default number of HDFS copies is 2. If the number of Core nodes in the cluster is greater than or equal to 3, the default number of HDFS copies is 3.

Figure 1-42 HDFS architecture



The HDFS component of MRS supports the following features:

- Supports erasure code, reducing data redundancy to 50% and improving reliability. In addition, the striped block storage structure is introduced to maximize the use of the capability of a single node and multiple disks in an existing cluster. After the coding process is introduced, the data write performance is improved, and the performance is close to that with the multi-copy redundancy.
- Supports balanced node scheduling on HDFS and balanced disk scheduling on a single node, improving HDFS storage performance after node or disk scale-out.

For details about the Hadoop architecture and principles, see **https://hadoop.apache.org/**.

1.3.7.2 HDFS HA Solution

HDFS HA Background

In versions earlier than Hadoop 2.0.0, SPOF occurs in the HDFS cluster. Each cluster has only one NameNode. If the host where the NameNode is located is faulty, the HDFS cluster cannot be used unless the NameNode is restarted or started on another host. This affects the overall availability of HDFS in the following aspects:

1. In the case of an unplanned event such as host breakdown, the cluster would be unavailable until the NameNode is restarted.
2. Planned maintenance tasks, such as software and hardware upgrade, will cause the cluster stop working.

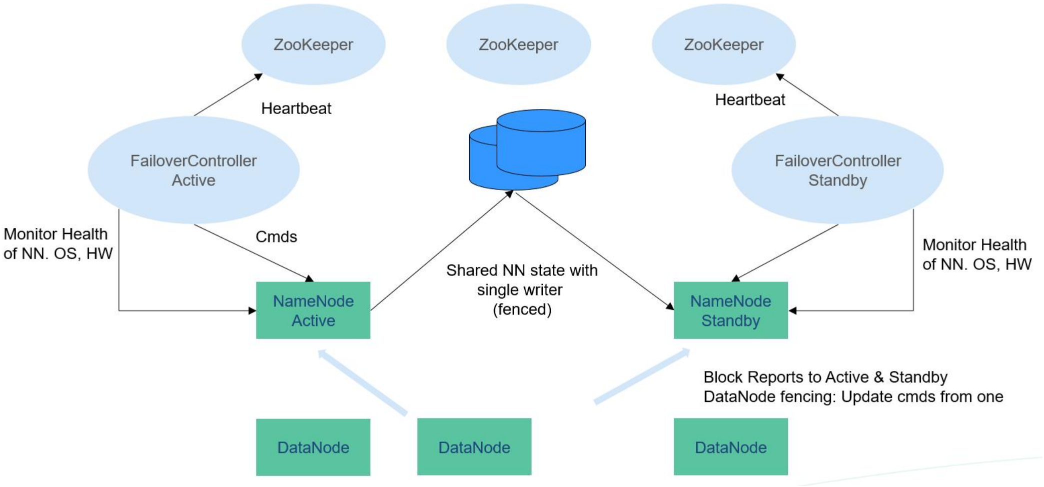
To solve the preceding problems, the HDFS HA solution enables a hot-swap NameNode backup for NameNodes in a cluster in automatic or manual (configurable) mode. When a machine fails (due to hardware failure), the active/ standby NameNode switches over automatically in a short time. When the active NameNode needs to be maintained, the administrator can manually perform an active/standby NameNode switchover to ensure cluster availability during maintenance.

For details about HDFS automatic failover, see

http://hadoop.apache.org/docs/r3.1.1/hadoop-project-dist/hadoop-hdfs/HDFSHighAvailabilityWithQJM.html#Automatic_Failover

HDFS HA Implementation

Figure 1-43 Typical HA deployment


ZooKeeper
ZooKeeper
ZooKeeper
Heartbeat
Heartbeat
FailoverController Active
FailoverController Standby
Monitor Health of NN. OS, HW
Cmds
Shared NN state with single writer (fenced)
Monitor Health of NN. OS, HW
NameNode Active
NameNode Standby
Block Reports to Active & Standby
DataNode fencing: Update cmds from one
DataNode
DataNode
DataNode

In a typical HA cluster (as shown in **Figure 1-43**), two NameNodes need to be configured on two independent servers, respectively. At any time point, one NameNode is in the active state, and the other NameNode is in the standby state. The active NameNode is responsible for all client operations in the cluster, while the standby NameNode maintains synchronization with the active node to provide fast switchover if necessary.

To keep the data synchronized with each other, both nodes communicate with a group of JournalNodes. When the active node modifies any file system's metadata, it will store the modification log to a majority of these JournalNodes. For example, if there are three JournalNodes, then the log will be saved on two of them at least. The standby node monitors changes of JournalNodes and synchronizes changes from the active node. Based on the modification log, the standby node applies the changes to the metadata of the local file system. Once a switchover occurs, the standby node can ensure its status is the same as that of the active node. This ensures that the metadata of the file system is synchronized between the active and standby nodes if the switchover is incurred by the failure of the active node.

To ensure fast switchover, the standby node needs to have the latest block information. Therefore, DataNodes send block information and heartbeat messages to two NameNodes at the same time.

It is vital for an HA cluster that only one of the NameNodes be active at any time. Otherwise, the namespace state would split into two parts, risking data loss or other incorrect results. To prevent the so-called "split-brain scenario", the JournalNodes will only ever allow a single NameNode to write data to it at a time. During switchover, the NameNode which is to become active will take over the role of writing data to JournalNodes. This effectively prevents the other NameNodes from being in the active state, allowing the new active node to safely proceed with switchover.

For more information about the HDFS HA solution, visit the following website:

http://hadoop.apache.org/docs/r3.1.1/hadoop-project-dist/hadoop-hdfs/HDFSHighAvailabilityWithQJM.html

1.3.7.3 Relationship Between HDFS and Other Components

Relationship Between HDFS and HBase

HDFS is a subproject of Apache Hadoop, which is used as the file storage system for HBase. HBase is located in the structured storage layer. HDFS provides highly reliable support for lower-layer storage of HBase. All the data files of HBase can be stored in the HDFS, except some log files generated by HBase.

Relationship Between HDFS and MapReduce

- HDFS features high fault tolerance and high throughput, and can be deployed on low-cost hardware for storing data of applications with massive data sets.
- MapReduce is a programming model used for parallel computation of large data sets (larger than 1 TB). Data computed by MapReduce comes from multiple data sources, such as Local FileSystem, HDFS, and databases. Most data comes from the HDFS. The high throughput of HDFS can be used to read massive data. After being computed, data can be stored in HDFS.

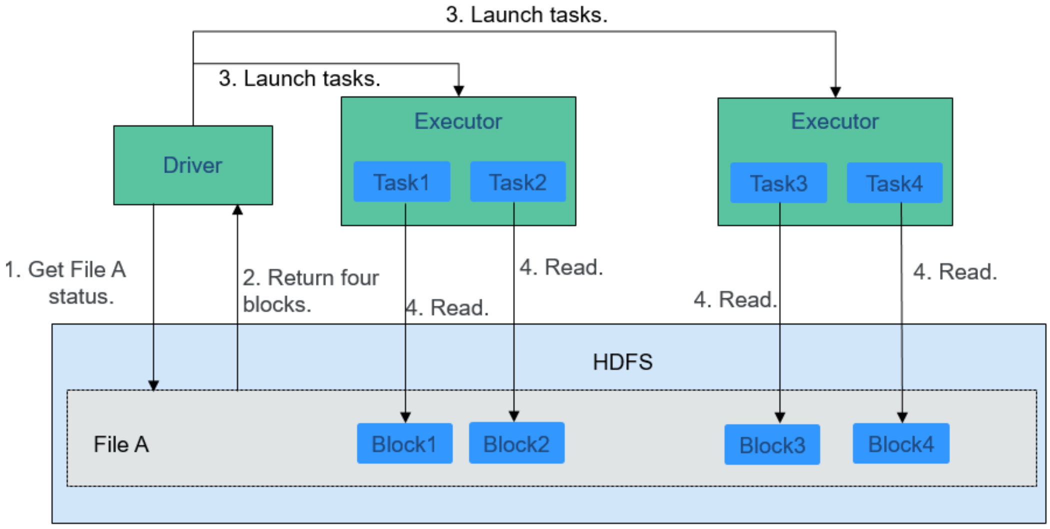
Relationship Between HDFS and Spark

Data computed by Spark comes from multiple data sources, such as local files and HDFS. Most data comes from HDFS which can read data in large scale for parallel computing. After being computed, data can be stored in HDFS.

Spark involves Driver and Executor. Driver schedules tasks and Executor runs tasks.

Figure 1-44 shows how data is read from a file.

Figure 1-44 File reading process

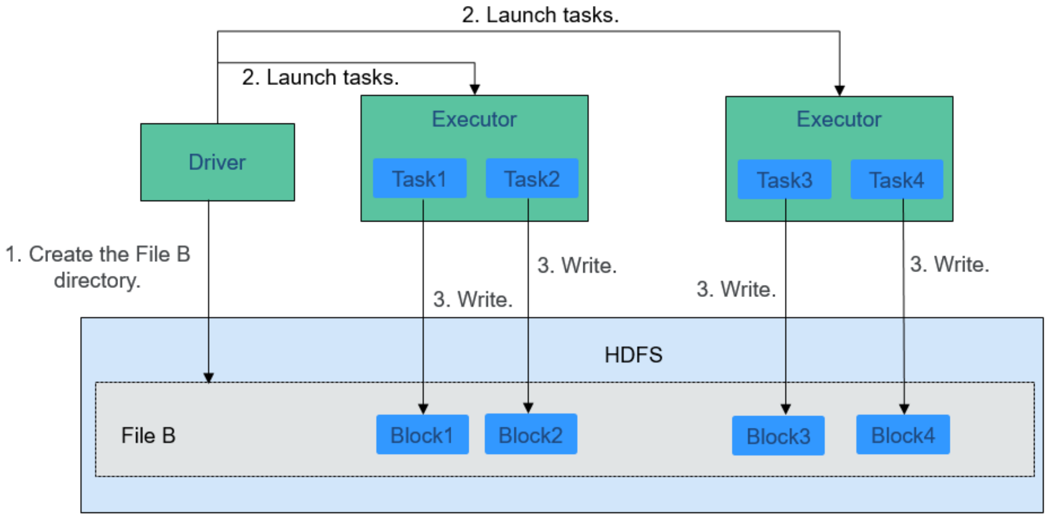


The file reading process is as follows:

1. Driver interconnects with HDFS to obtain the information of File A.
2. The HDFS returns the detailed block information about this file.
3. Driver sets a parallel degree based on the block data amount, and creates multiple tasks to read the blocks of this file.
4. Executor runs the tasks and reads the detailed blocks as part of the Resilient Distributed Dataset (RDD).

Figure 1-45 shows how data is written to a file.

Figure 1-45 File writing process



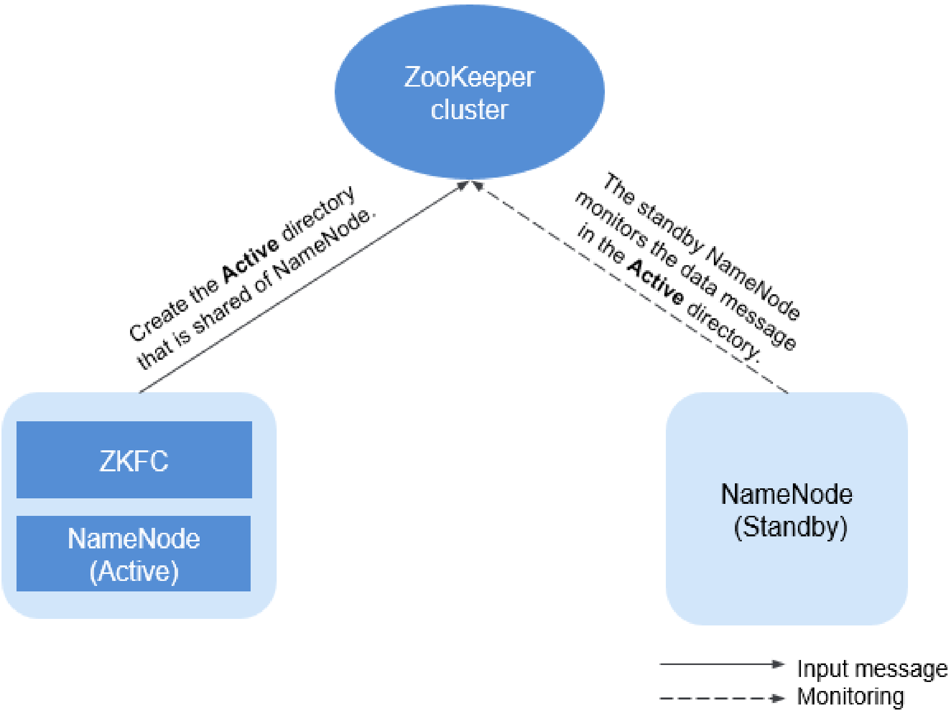
The file writing process is as follows:

1. Driver creates a directory where the file is to be written.
2. Based on the RDD distribution status, the number of tasks related to data writing is computed, and these tasks are sent to Executor.
3. Executor runs these tasks, and writes the computed RDD data to the directory created in 1.

Relationship Between HDFS and ZooKeeper

Figure 1-46 shows the relationship between ZooKeeper and HDFS.

Figure 1-46 Relationship between ZooKeeper and HDFS


ZooKeeper cluster
Create the **Active** directory that is shared of NameNode.
The standby NameNode monitors the data message in the **Active** directory.
ZKFC
NameNode (Active)
NameNode (Standby)
Input message
Monitoring

As the client of a ZooKeeper cluster, ZKFailoverController (ZKFC) monitors the status of NameNode. ZKFC is deployed only in the node where NameNode resides, and in both the active and standby HDFS NameNodes.

1. The ZKFC connects to ZooKeeper and saves information such as host names to ZooKeeper under the znode directory **/hadoop-ha**. NameNode that creates the directory first is considered as the active node, and the other is the standby node. NameNodes read the NameNode information periodically through ZooKeeper.
2. When the process of the active node ends abnormally, the standby NameNode detects changes in the **/hadoop-ha** directory through ZooKeeper, and then takes over the service of the active NameNode.

1.3.7.4 HDFS Enhanced Open Source Features

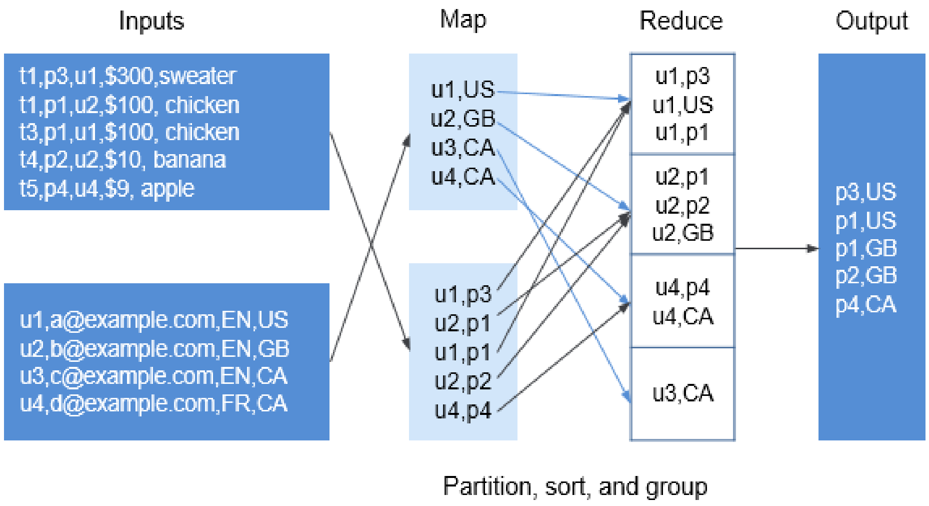
Enhanced Open Source Feature: File Block Colocation

In the offline data summary and statistics scenario, Join is a frequently used computing function, and is implemented in MapReduce as follows:

1. The Map task processes the records in the two table files into Join Key and Value, performs hash partitioning by Join Key, and sends the data to different Reduce tasks for processing.
2. Reduce tasks read data in the left table recursively in the nested loop mode and traverse each line of the right table. If join key values are identical, join results are output.

   The preceding method sharply reduces the performance of the join calculation. Because a large amount of network data transfer is required when the data stored in different nodes is sent from MAP to Reduce, as shown in **Figure 1-47**.

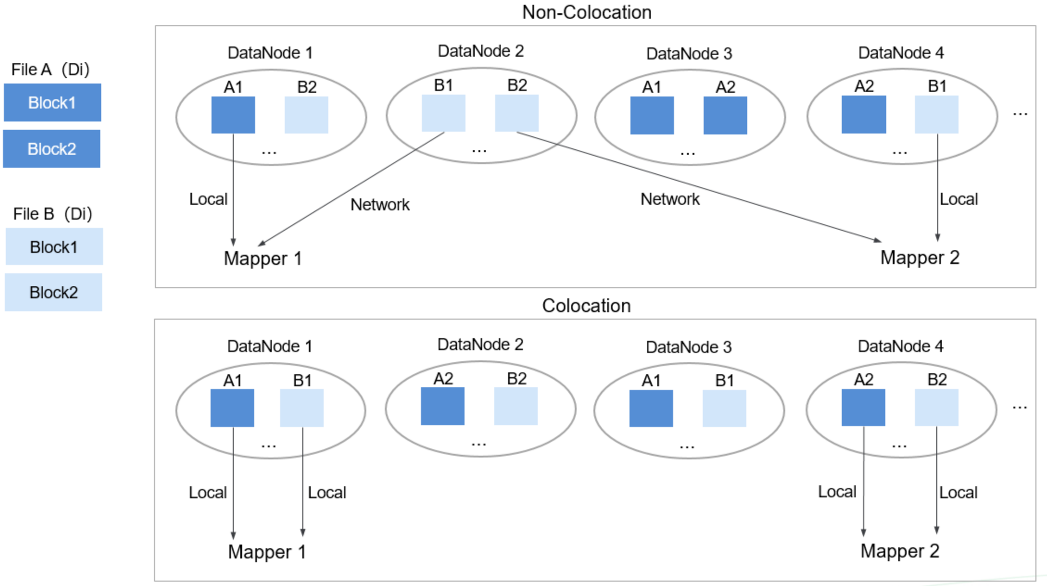
Figure 1-47 Data transmission in the non-colocation scenario



Data tables are stored in physical file system by HDFS block. Therefore, if two to-be-joined blocks are put into the same host accordingly after they are partitioned by join key, you can obtain the results directly from Map join in the local node without any data transfer in the Reduce process of the join calculation. This will greatly improve the performance.

With the identical distribution feature of HDFS data, a same distribution ID is allocated to files, FileA and FileB, on which association and summation calculations need to be performed. In this way, all the blocks are distributed together, and calculation can be performed without retrieving data across nodes, which greatly improves the MapReduce join performance.

Figure 1-48 Data block distribution in colocation and non-colocation scenarios

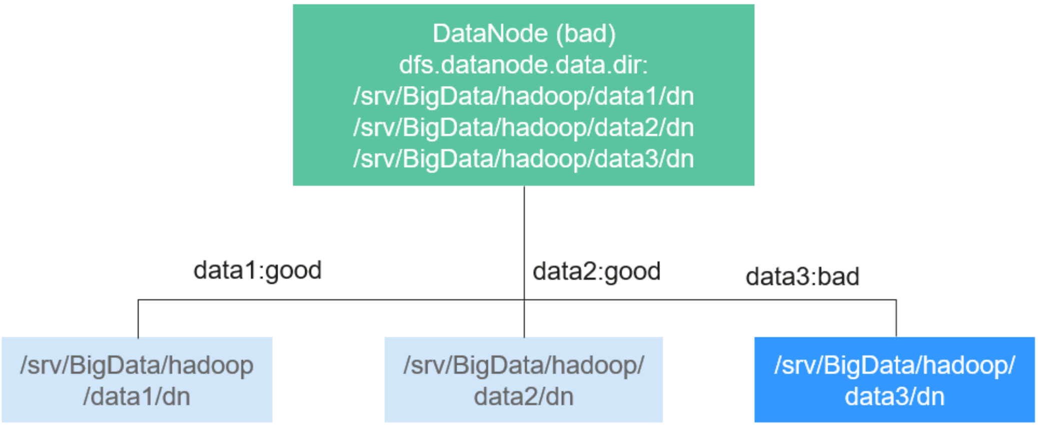


Enhanced Open Source Feature: Damaged Hard Disk Volume Configuration

In the open source version, if multiple data storage volumes are configured for a DataNode, the DataNode stops providing services by default if one of the volumes is damaged. If the configuration item **dfs.datanode.failed.volumes.tolerated** is set to specify the number of damaged volumes that are allowed, DataNode continues to provide services when the number of damaged volumes does not exceed the threshold.

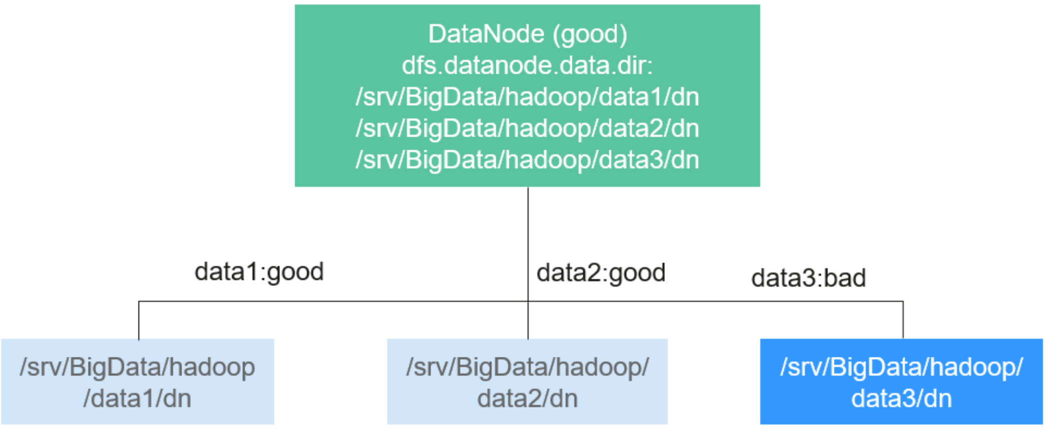
The value of **dfs.datanode.failed.volumes.tolerated** ranges from -1 to the number of disk volumes configured on the DataNode. The default value is **-1**, as shown in **Figure 1-49**.

Figure 1-49 Item being set to 0


DataNode (bad)
dfs.datanode.data.dir:
/srv/BigData/hadoop/data1/dn
/srv/BigData/hadoop/data2/dn
/srv/BigData/hadoop/data3/dn
data1:good
data2:good
data3:bad
/srv/BigData/hadoop/data1/dn
/srv/BigData/hadoop/data2/dn
/srv/BigData/hadoop/data3/dn

For example, three data storage volumes are mounted to a DataNode, and **dfs.datanode.failed.volumes.tolerated** is set to 1. In this case, if one data storage volume of the DataNode is unavailable, this DataNode can still provide services, as shown in **Figure 1-50**.

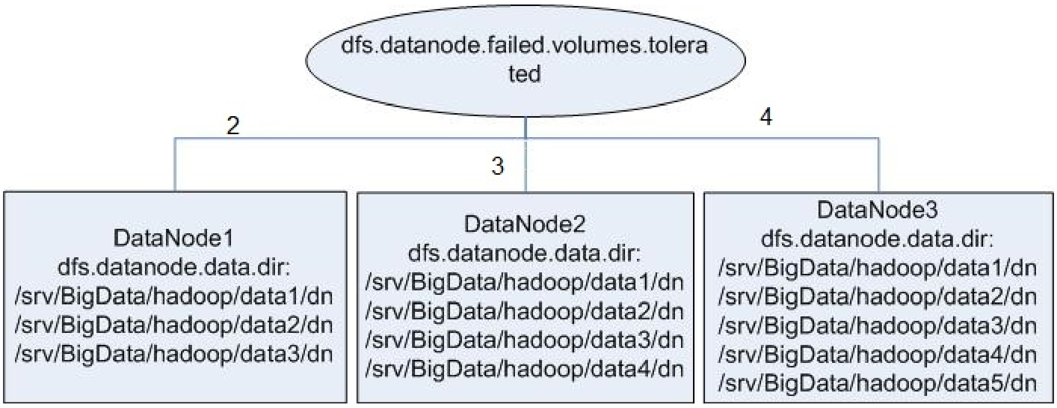
Figure 1-50 Item being set to 1


DataNode (good)
dfs.datanode.data.dir:
/srv/BigData/hadoop/data1/dn
/srv/BigData/hadoop/data2/dn
/srv/BigData/hadoop/data3/dn
data1:good
data2:good
data3:bad
/srv/BigData/hadoop/data1/dn
/srv/BigData/hadoop/data2/dn
/srv/BigData/hadoop/data3/dn

This native configuration item has some defects. When the number of data storage volumes in each DataNode is inconsistent, you need to configure each DataNode independently instead of generating the unified configuration file for all nodes.

Assume that there are three DataNodes in a cluster. The first node has three data directories, the second node has four, and the third node has five. If you want to ensure that DataNode services are available when only one data directory is available, you need to perform the configuration as shown in **Figure 1-51**.

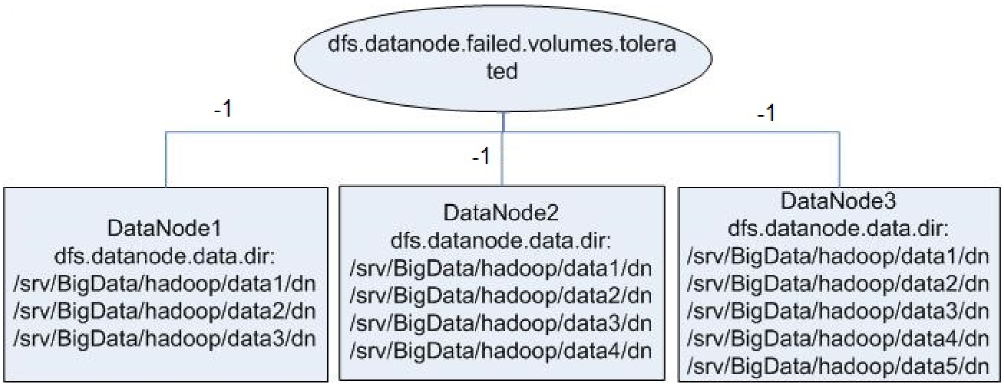
Figure 1-51 Attribute configuration before being enhanced


dfs.datanode.failed.volumes.tolerated
2
3
4
DataNode1
dfs.datanode.data.dir:
/srv/BigData/hadoop/data1/dn
/srv/BigData/hadoop/data2/dn
/srv/BigData/hadoop/data3/dn
DataNode2
dfs.datanode.data.dir:
/srv/BigData/hadoop/data1/dn
/srv/BigData/hadoop/data2/dn
/srv/BigData/hadoop/data3/dn
/srv/BigData/hadoop/data4/dn
DataNode3
dfs.datanode.data.dir:
/srv/BigData/hadoop/data1/dn
/srv/BigData/hadoop/data2/dn
/srv/BigData/hadoop/data3/dn
/srv/BigData/hadoop/data4/dn
/srv/BigData/hadoop/data5/dn

In self-developed enhanced HDFS, this configuration item is enhanced, with a value **-1** added. When this configuration item is set to **-1**, all DataNodes can provide services as long as one data storage volume in all DataNodes is available.

To resolve the problem in the preceding example, set this configuration to **-1**, as shown in **Figure 1-52**.

Figure 1-52 Attribute configuration after being enhanced


dfs.datanode.failed.volumes.tolera
ted
-1
-1
-1
DataNode1
dfs.datanode.data.dir:
/srv/BigData/hadoop/data1/dn
/srv/BigData/hadoop/data2/dn
/srv/BigData/hadoop/data3/dn
DataNode2
dfs.datanode.data.dir:
/srv/BigData/hadoop/data1/dn
/srv/BigData/hadoop/data2/dn
/srv/BigData/hadoop/data3/dn
/srv/BigData/hadoop/data4/dn
DataNode3
dfs.datanode.data.dir:
/srv/BigData/hadoop/data1/dn
/srv/BigData/hadoop/data2/dn
/srv/BigData/hadoop/data3/dn
/srv/BigData/hadoop/data4/dn
/srv/BigData/hadoop/data5/dn

Enhanced Open Source Feature: HDFS Startup Acceleration

In HDFS, when NameNodes start, the metadata file FsImage needs to be loaded. Then, DataNodes will report the data block information after the DataNodes startup. When the data block information reported by DataNodes reaches the preset percentage, NameNodes exits safe mode to complete the startup process. If the number of files stored on the HDFS reaches the million or billion level, the two processes are time-consuming and will lead to a long startup time of the NameNode. Therefore, this version optimizes the process of loading metadata file FsImage.

In the open source HDFS, FsImage stores all types of metadata information. Each type of metadata information (such as file metadata information and folder metadata information) is stored in a section block, respectively. These section blocks are loaded in serial mode during startup. If a large number of files and folders are stored on the HDFS, loading of the two sections is time-consuming, prolonging the HDFS startup time. HDFS NameNode divides each type of metadata by segments and stores the data in multiple sections when generating the FsImage files. When the NameNodes start, sections are loaded in parallel mode. This accelerates the HDFS startup.

Enhanced Open Source Feature: Label-based Block Placement Policies (HDFS Nodelabel)

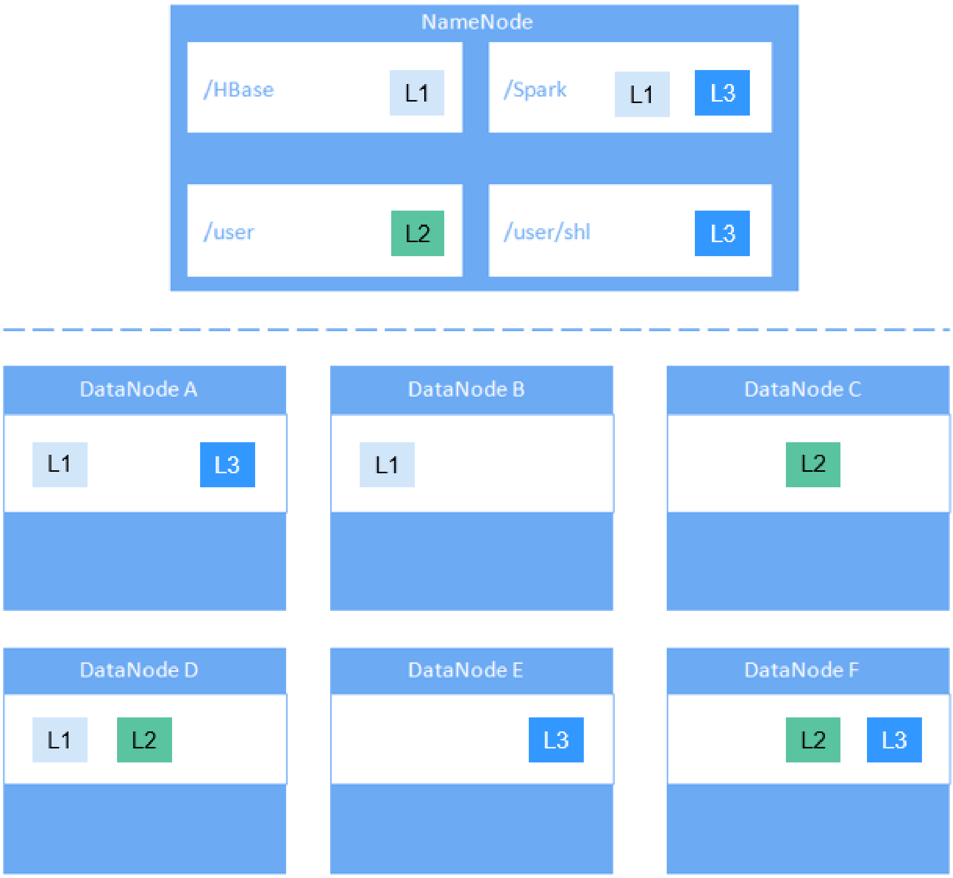
You need to configure the nodes for storing HDFS file data blocks based on data features. You can configure a label expression to an HDFS directory or file and assign one or more labels to a DataNode so that file data blocks can be stored on specified DataNodes. If the label-based data block placement policy is used for selecting DataNodes to store the specified files, the DataNode range is specified based on the label expression. Then proper nodes are selected from the specified range.

- You can store the replicas of data blocks to the nodes with different labels accordingly. For example, store two replicas of the data block to the node labeled with L1, and store other replicas of the data block to the nodes labeled with L2.
- You can set the policy in case of block placement failure, for example, select a node from all nodes randomly.

Figure 1-53 gives an example:

- Data in **/HBase** is stored in A, B, and D.
- Data in **/Spark** is stored in A, B, D, E, and F.
- Data in **/user** is stored in C, D, and F.
- Data in **/user/shl** is stored in A, E, and F.

Figure 1-53 Example of label-based block placement policy



Enhanced Open Source Feature: HDFS Load Balance

The current read and write policies of HDFS are mainly for local optimization without considering the actual load of nodes or disks. Based on I/O loads of different nodes, the load balance of HDFS ensures that when read and write operations are performed on the HDFS client, the node with low I/O load is selected to perform such operations to balance I/O load and fully utilize the overall throughput of the cluster.

If HDFS Load Balance is enabled during file writing, the NameNode selects a DataNode (in the order of local node, local rack, and remote rack). If the I/O load of the selected node is heavy, the NameNode will choose another DataNode with lighter load.

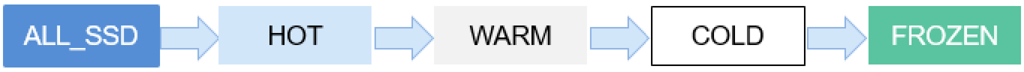
If HDFS Load Balance is enabled during file reading, an HDFS client sends a request to the NameNode to provide the list of DataNodes that store the block to be read. The NameNode returns a list of DataNodes sorted by distance in the network topology. With the HDFS Load Balance feature, the DataNodes on the list are also sorted by their I/O load. The DataNodes with heavy load are at the bottom of the list.

Enhanced Open Source Feature: HDFS Auto Data Movement

Hadoop has been used for batch processing of immense data in a long time. The existing HDFS model is used to fit the needs of batch processing applications very well because such applications focus more on throughput than delay.

However, as Hadoop is increasingly used for upper-layer applications that demand frequent random I/O access such as Hive and HBase, low latency disks such as solid state disk (SSD) are favored in delay-sensitive scenarios. To cater to the trend, HDFS supports a variety of storage types. Users can choose a storage type according to their needs.

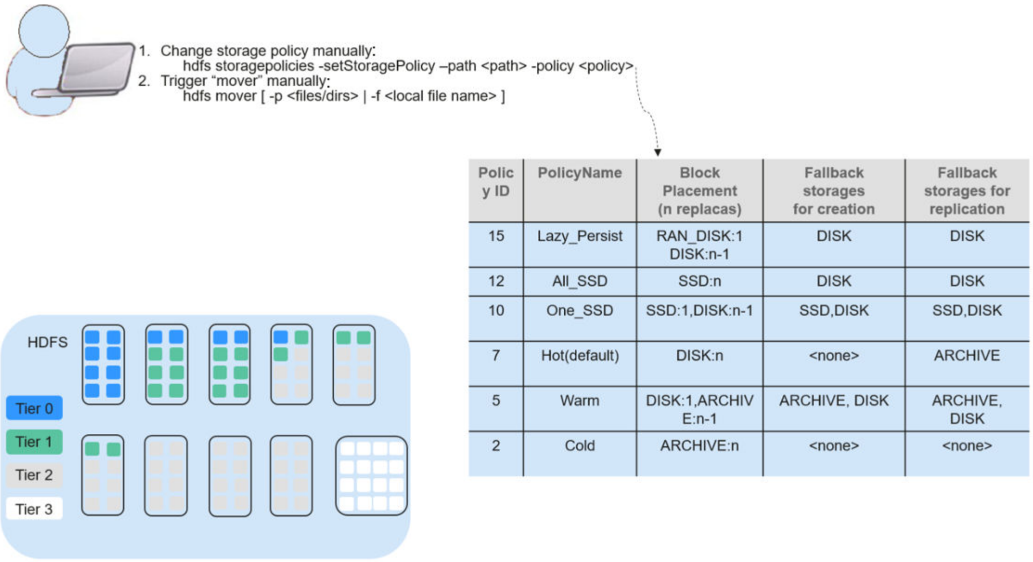
Storage policies vary depending on how frequently data is used. For example, if data that is frequently accessed in the HDFS is marked as **ALL_SSD** or **HOT**, the data that is accessed several times may be marked as **WARM**, and data that is rarely accessed (only once or twice access) can be marked as **COLD**. You can select different data storage policies based on the data access frequency.



However, low latency disks are far more expensive than spinning disks. Data typically sees heavy initial usage with decline in usage over a period of time. Therefore, it can be useful if data that is no longer used is moved out from expensive disks to cheaper ones storage media.

A typical example is storage of detail records. New detail records are imported into SSD because they are frequently queried by upper-layer applications. As access frequency to these detail records declines, they are moved to cheaper storage.

Before automatic data movement is achieved, you have to manually determine by service type whether data is frequently used, manually set a data storage policy, and manually trigger the HDFS Auto Data Movement Tool, as shown in the figure below.



1. Change storage policy manually:
   hdfs storagepolicies -setStoragePolicy –path <path> -policy <policy>
2. Trigger “mover” manually:
   hdfs mover [-p <files/dirs> | -f <local file name>]

Polic y ID	PolicyName	Block Placement (n replacas)	Fallback storages for creation	Fallback storages for replication
15	Lazy_Persist	RAN_DISK:1 DISK:n-1	DISK	DISK
12	All_SSD	SSD:n	DISK	DISK
10	One_SSD	SSD:1,DISK:n-1	SSD,DISK	SSD,DISK
7	Hot(default)	DISK:n	<none>	ARCHIVE
5	Warm	DISK:1,ARCHIVE:n-1	ARCHIVE, DISK	ARCHIVE, DISK
2	Cold	ARCHIVE:n	<none>	<none>

If aged data can be automatically identified and moved to cheaper storage (such as disk/archive), you will see significant cost cuts and data management efficiency improvement.

The HDFS Auto Data Movement Tool is at the core of HDFS Auto Data Movement. It automatically sets a storage policy depending on how frequently data is used. Specifically, functions of the HDFS Auto Data Movement Tool can:

- Mark a data storage policy as **All_SSD**, **One_SSD**, **Hot**, **Warm**, **Cold**, or **FROZEN** according to age, access time, and manual data movement rules.
- Define rules for distinguishing cold and hot data based on the data age, access time, and manual migration rules.
- Define the action to be taken if age-based rules are met.

  MARK: the action for identifying whether data is frequently or rarely used based on the age rules and setting a data storage policy. **MOVE**: the action for invoking the HDFS Auto Data Movement Tool and moving data based on the age rules to identify whether data is frequently or rarely used after you have determined the corresponding storage policy.

  - **MARK**: identifies whether data is frequently or rarely used and sets the data storage policy.
  - **MOVE**: the action for invoking the HDFS Auto Data Movement Tool and moving data across tiers.
  - **SET_REPL**: the action for setting new replica quantity for a file.
  - **MOVE_TO_FOLDER**: the action for moving files to a target folder.
  - **DELETE**: the action for deleting a file or directory.
  - **SET_NODE_LABEL**: the action for setting node labels of a file.

With the HDFS Auto Data Movement feature, you only need to define age based on access time rules. HDFS Auto Data Movement Tool matches data according to age-based rules, sets storage policies, and moves data. In this way, data management efficiency and cluster resource efficiency are improved.

1.3.8 Hive

1.3.8.1 Hive Basic Principles

Hive is a data warehouse infrastructure built on Hadoop. It provides a series of tools that can be used to extract, transform, and load (ETL) data. Hive is a mechanism that can store, query, and analyze mass data stored on Hadoop. Hive defines simple SQL-like query language, which is known as HiveQL. It allows a user familiar with SQL to query data. Hive data computing depends on MapReduce, Spark, and Tez.

The new execution engine **Tez** is used to replace the original MapReduce, greatly improving performance. Tez can convert multiple dependent jobs into one job, so only once HDFS write is required and fewer transit nodes are needed, greatly improving the performance of DAG jobs.

Hive provides the following functions:

- Analyzes massive structured data and summarizes analysis results.

- Allows complex MapReduce jobs to be compiled in SQL languages.
- Supports flexible data storage formats, including JavaScript object notation (JSON), comma separated values (CSV), TextFile, RCFile, SequenceFile, and ORC (Optimized Row Columnar).

Hive system structure:

- User interface: Three user interfaces are available, that is, CLI, Client, and WUI. CLI is the most frequently-used user interface. A Hive transcript is started when CLI is started. Client refers to a Hive client, and a client user connects to the Hive Server. When entering the client mode, you need to specify the node where the Hive Server resides and start the Hive Server on this node. The web UI is used to access Hive through a browser. MRS can access Hive only in client mode.
- Metadata storage: Hive stores metadata into databases, for example, MySQL and Derby. Metadata in Hive includes a table name, table columns and partitions and their properties, table properties (indicating whether a table is an external table), and the directory where table data is stored.

Hive Framework

Hive is a single-instance service process that provides services by translating HQL into related MapReduce jobs or HDFS operations. **Figure 1-54** shows how Hive is connected to other components.

Figure 1-54 Hive framework

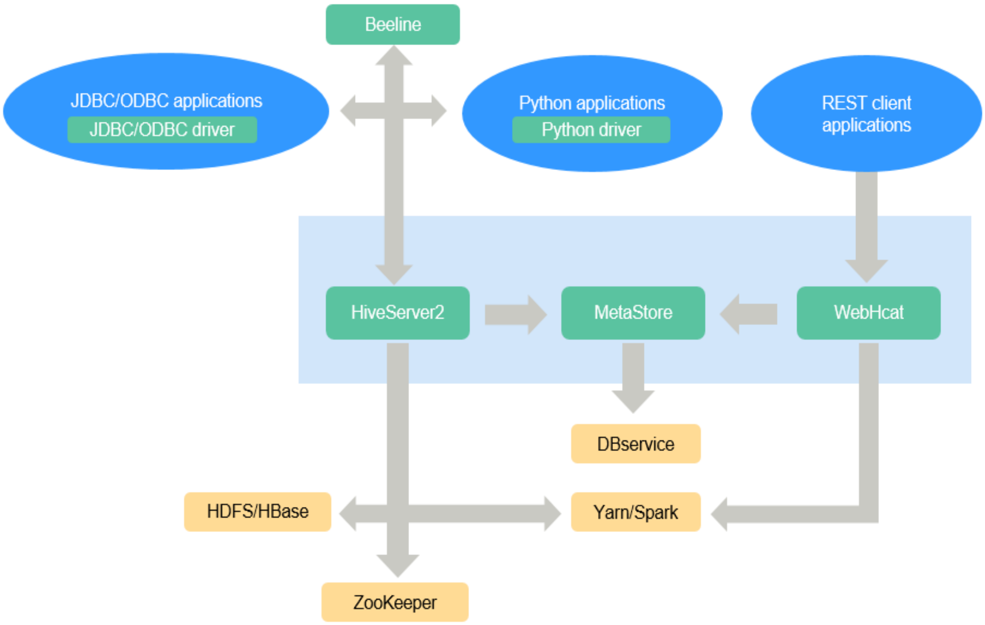
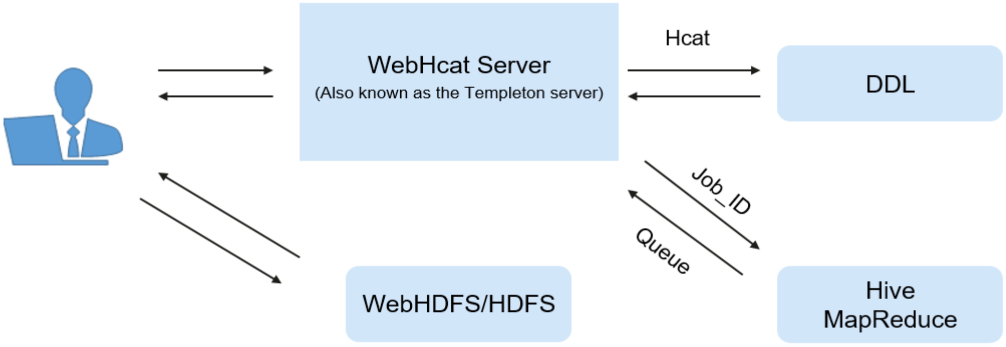


Table 1-9 Module description

Module	Description
HiveServer	Multiple HiveServers can be deployed in a cluster to share loads. HiveServer provides Hive database services externally, translates HQL statements into related YARN tasks or HDFS operations to complete data extraction, conversion, and analysis.
MetaStore	● Multiple MetaStores can be deployed in a cluster to share loads. MetaStore provides Hive metadata services as well as reads, writes, maintains, and modifies the structure and properties of Hive tables. ● MetaStore provides Thrift APIs for HiveServer, Spark, WebHCat, and other MetaStore clients to access and operate metadata.
WebHCat	Multiple WebHCats can be deployed in a cluster to share loads. WebHCat provides REST APIs and runs the Hive commands through the REST APIs to submit MapReduce jobs.
Hive client	Hive client includes the human-machine command-line interface (CLI) Beeline, JDBC drive for JDBC applications, Python driver for Python applications, and HCatalog JAR files for MapReduce.
ZooKeeper cluster	As a temporary node, ZooKeeper records the IP address list of each HiveServer instance. The client driver connects to ZooKeeper to obtain the list and selects corresponding HiveServer instances based on the routing mechanism.
HDFS/HBase cluster	The HDFS cluster stores the Hive table data.
MapReduce/YARN cluster	Provides distributed computing services. Most Hive data operations rely on MapReduce. The main function of HiveServer is to translate HQL statements into MapReduce jobs to process massive data.

HCatalog is built on Hive Metastore and incorporates the DDL capability of Hive. HCatalog is also a Hadoop-based table and storage management layer that enables convenient data read/write on tables of HDFS by using different data processing tools such as Pig and MapReduce. Besides, HCatalog also provides read/write APIs for these tools and uses a Hive CLI to publish commands for defining data and querying metadata. After encapsulating these commands, WebHCat Server can provide RESTful APIs, as shown in **Figure 1-55**.

Figure 1-55 WebHCat logical architecture



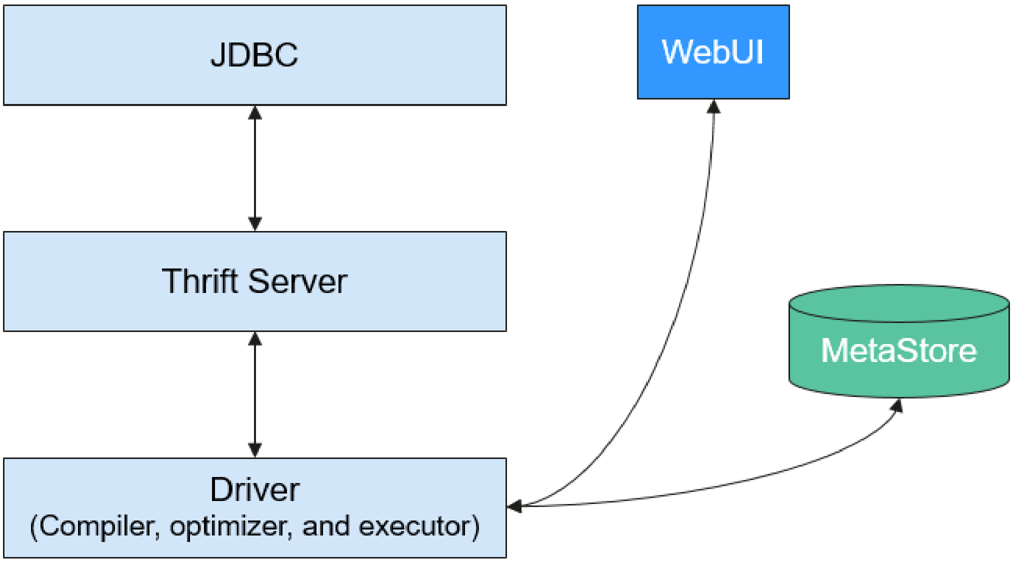
Principles

Hive functions as a data warehouse based on HDFS and MapReduce architecture and translates HQL statements into MapReduce jobs or HDFS operations. For details about Hive and HQL, see **HiveQL Language Manual**.

Figure 1-56 shows the Hive structure.

- **Metastore**: reads, writes, and updates metadata such as tables, columns, and partitions. Its lower layer is relational databases.
- **Driver**: manages the lifecycle of HiveQL execution and participates in the entire Hive job execution.
- **Compiler**: translates HQL statements into a series of interdependent Map or Reduce jobs.
- **Optimizer**: is classified into logical optimizer and physical optimizer to optimize HQL execution plans and MapReduce jobs, respectively.
- **Executor**: runs Map or Reduce jobs based on job dependencies.
- **ThriftServer**: functions as the servers of JDBC, provides Thrift APIs, and integrates with Hive and other applications.
- **Clients**: include the WebUI and JDBC APIs and provides APIs for user access.

Figure 1-56 Hive framework



1.3.8.2 Hive CBO Principles

Hive CBO Principles

CBO is short for Cost-Based Optimization.

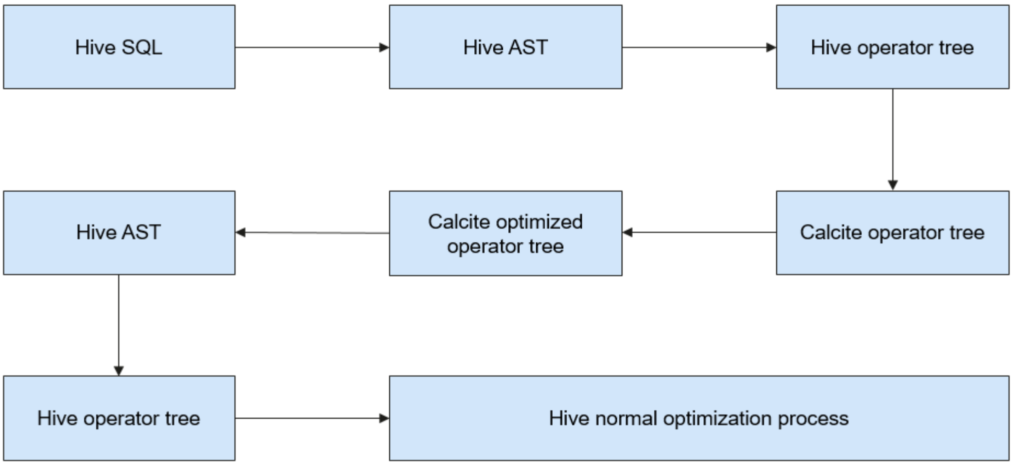
It will optimize the following:

During compilation, the CBO calculates the most efficient join sequence based on tables and query conditions involved in query statements to reduce time and resources required for query.

In Hive, the CBO is implemented as follows:

Hive uses open-source component Apache Calcite to implement the CBO. SQL statements are first converted into Hive Abstract Syntax Trees (ASTs) and then into RelNodes that can be identified by Calcite. After Calcite adjusts the join sequence in RelNodes, RelNodes are converted into ASTs by Hive to continue the logical and physical optimization. **Figure 1-57** shows the working flow.

Figure 1-57 CBO Implementation process



Calcite adjusts the join sequence as follows:

1. A table is selected as the first table from the tables to be joined.
2. The second and third tables are selected based on the cost. In this way, multiple different execution plans are obtained.
3. A plan with the minimum costs is calculated and serves as the final sequence.

The cost calculation method is as follows:

In the current version, costs are measured based on the number of data entries after joining. Fewer data entries mean less cost. The number of joined data entries depends on the selection rate of joined tables. The number of data entries in a table is obtained based on the table-level statistics.

The number of data entries in a table after filtering is estimated based on the column-level statistics, including the maximum values (max), minimum values (min), and Number of Distinct Values (NDV).

For example, there is a table **table_a** whose total number of data records is 1,000,000 and NDV is 50. The query conditions are as follows:

```
Select * from table_a where colum_a='value1';
```

The estimated number of queried data entries is: 1,000,000 x 1/50 = 20,000. The selection rate is 2%.

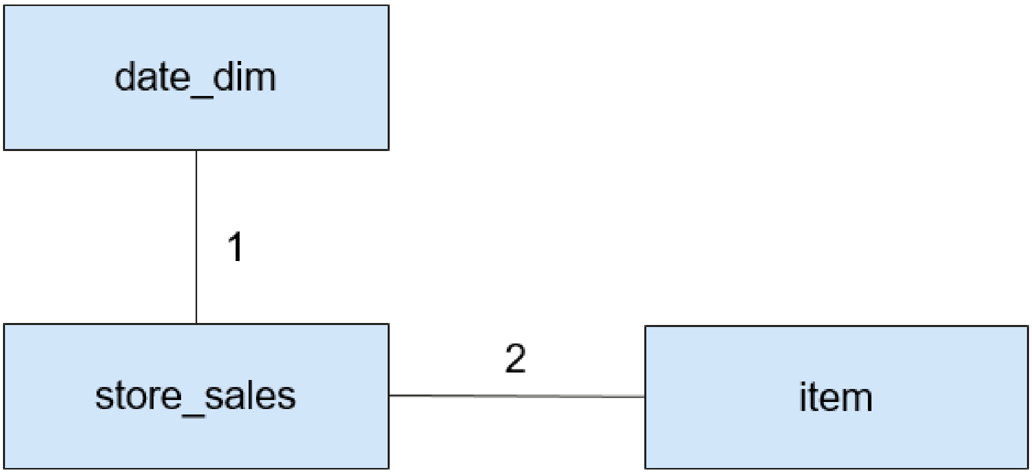
The following takes the TPC-DS Q3 as an example to describe how the CBO adjusts the join sequence:

```
select
    dt.d_year,
    item.i_brand_id brand_id,
    item.i_brand brand,
    sum(ss_ext_sales_price) sum_agg
from
    date_dim dt,
    store_sales,
    item
where
    dt.d_date_sk = store_sales.ss_sold_date_sk
    and store_sales.ss_item_sk = item.i_item_sk
```

```
   and item.i_manufact_id = 436
   and dt.d_moy = 12
group by dt.d_year , item.i_brand , item.i_brand_id
order by dt.d_year , sum_agg desc , brand_id
limit 10;
```

Statement explanation: This statement indicates that inner join is performed for three tables: table **store_sales** is a fact table with about 2,900,000,000 data entries, table **date_dim** is a dimension table with about 73,000 data entries, and table **item** is a dimension table with about 18,000 data entries. Each table has filtering conditions. **Figure 1-58** shows the join relationship.

Figure 1-58 Join relationship



The CBO must first select the tables that bring the best filtering effect for joining.

By analyzing min, max, NDV, and the number of data entries, the CBO estimates the selection rates of different dimension tables, as shown in **Table 1-10**.

Table 1-10 Data filtering

Table	Number of Original Data Entries	Number of Data Entries After Filtering	Selection Rate
date_dim	73,000	6,200	8.5%
item	18,000	19	0.1%

The selection rate can be estimated as follows: Selection rate = Number of data entries after filtering/Number of original data entries

As shown in the preceding table, the **item** table has a better filtering effect. Therefore, the CBO joins the **item** table first before joining the **date_dim** table.

Figure 1-59 shows the join process when the CBO is disabled.

Figure 1-59 Join process when the CBO is disabled

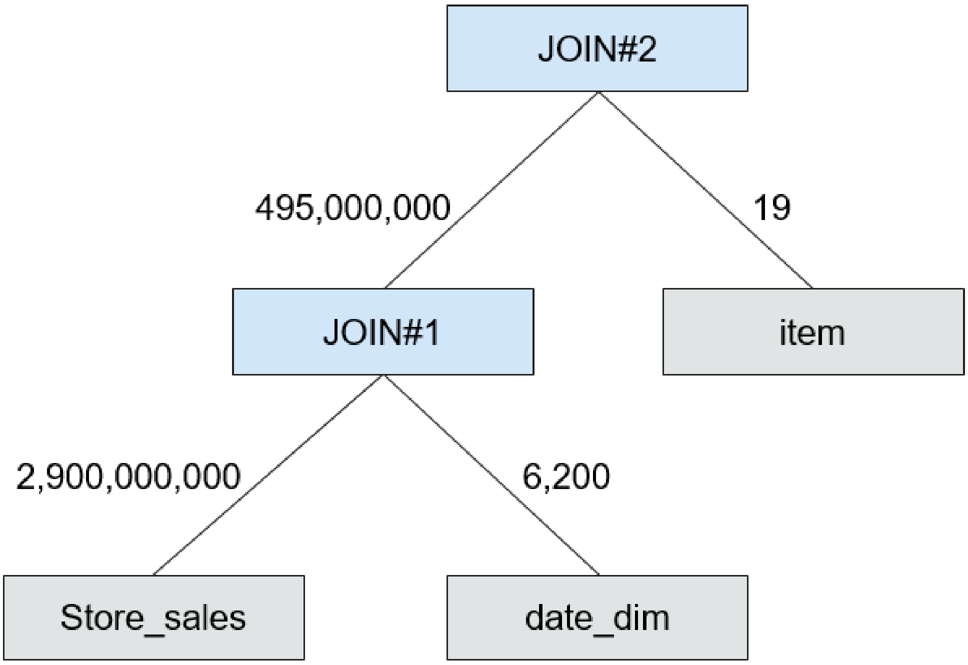
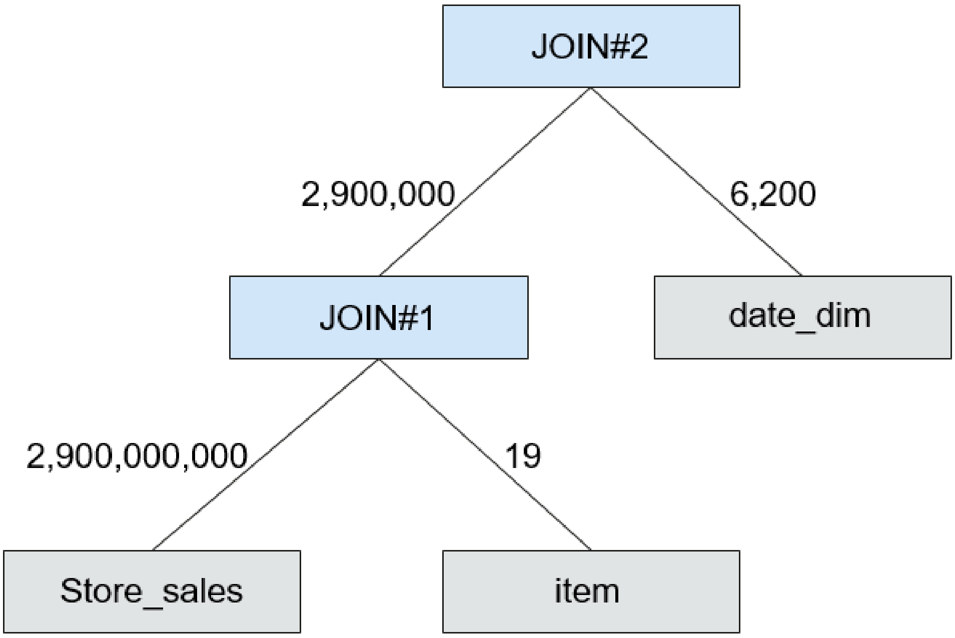


Figure 1-60 shows the join process when the CBO is enabled.

Figure 1-60 Join process when the CBO is enabled



After the CBO is enabled, the number of intermediate data entries is reduced from 495,000,000 to 2,900,000 and thus the execution time can be remarkably reduced.

1.3.8.3 Relationship Between Hive and Other Components

Relationship Between Hive and HDFS

Hive is a sub-project of Apache Hadoop, which uses HDFS as the file storage system. It parses and processes structured data with highly reliable underlying storage supported by HDFS. All data files in the Hive database are stored in HDFS, and all data operations on Hive are also performed using HDFS APIs.

Relationship Between Hive and MapReduce

Hive data computing depends on MapReduce. MapReduce is also a sub-project of Apache Hadoop and is a parallel computing framework based on HDFS. During data analysis, Hive parses HQL statements submitted by users into MapReduce tasks and submits the tasks for MapReduce to execute.

Relationship Between Hive and Tez

Tez, an open-source project of Apache, is a distributed computing framework that supports directed acyclic graphs (DAGs). When Hive uses the Tez engine to analyze data, it parses HQL statements submitted by users into Tez tasks and submits the tasks to Tez for execution.

Relationship Between Hive and DBService

MetaStore (metadata service) of Hive processes the structure and attribute information of Hive metadata, such as Hive databases, tables, and partitions. The information needs to be stored in a relational database and is managed and processed by MetaStore. In the product, the metadata of Hive is stored and maintained by the DBService component, and the metadata service is provided by the Metadata component.

1.3.8.4 Enhanced Open Source Feature

Enhanced Open Source Feature: HDFS Colocation

HDFS Colocation is the data location control function provided by HDFS. The HDFS Colocation API stores associated data or data on which associated operations are performed on the same storage node.

Hive supports HDFS Colocation. When Hive tables are created, after the locator information is set for table files, the data files of related tables are stored on the same storage node. This ensures convenient and efficient data computing among associated tables.

Enhanced Open Source Feature: Column Encryption

Hive supports encryption of one or more columns. The columns to be encrypted and the encryption algorithm can be specified when a Hive table is created. When data is inserted into the table using the INSERT statement, the related columns are encrypted. The Hive column encryption does not support views and the Hive over HBase scenario.

The Hive column encryption mechanism supports two encryption algorithms that can be selected to meet site requirements during table creation:

- AES (the encryption class is **org.apache.hadoop.hive.serde2.AESRewriter**)
- SMS4 (the encryption class is **org.apache.hadoop.hive.serde2.SMS4Rewriter**)

Enhanced Open Source Feature: HBase Deletion

Due to the limitations of underlying storage systems, Hive does not support the ability to delete a single piece of table data. In Hive on HBase, Hive in the MRS solution supports the ability to delete a single piece of HBase table data. Using a specific syntax, Hive can delete one or more pieces of data from an HBase table.

Enhanced Open Source Feature: Row Delimiter

In most cases, a carriage return character is used as the row delimiter in Hive tables stored in text files, that is, the carriage return character is used as the terminator of a row during queries.

However, some data files are delimited by special characters, and not a carriage return character.

MRS Hive allows you to specify different characters or character combinations as row delimiters for Hive data in text files.

Enhanced Open Source Feature: HTTPS/HTTP-based REST API Switchover

WebHCat provides external REST APIs for Hive. By default, the open source community version uses the HTTP protocol.

MRS Hive supports the HTTPS protocol that is more secure, and enables switchover between the HTTP protocol and the HTTPS protocol.

Enhanced Open Source Feature: Transform Function

The Transform function is not allowed by Hive of the open source version. MRS Hive supports the configuration of the Transform function. The function is disabled by default, which is the same as that of the open source community version.

Users can modify configurations of the Transform function to enable the function. However, security risks exist when the Transform function is enabled.

Enhanced Open Source Feature: Temporary Function Creation Without ADMIN Permission

You must have **ADMIN** permission when creating temporary functions on Hive of the open source community version. MRS Hive supports the configuration of the function for creating temporary functions with **ADMIN** permission. The function is disabled by default, which is the same as that of the open-source community version.

You can modify configurations of this function. After the function is enabled, you can create temporary functions without **ADMIN** permission.

Enhanced Open Source Feature: Database Authorization

In the Hive open source community version, only the database owner can create tables in the database. You can be granted with the **CREATE** and **SELECT** permissions on tables by MRS Hive in a database. After you are granted with the permission to query data in the database, the system automatically associates the query permission on all tables in the database.

Enhanced Open Source Feature: Column Authorization

The Hive open source community version supports only table-level permission control. MRS Hive supports column-level permission control. You can be granted with column-level permissions, such as **SELECT**, **INSERT**, and **UPDATE**.

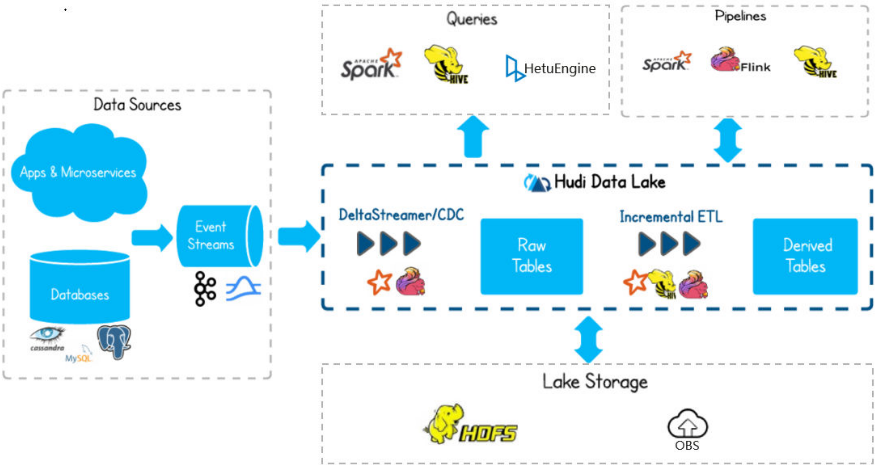
1.3.8.5 Hudi

Hudi is a data lake table format that provides the ability to update and delete data as well as consume new data on HDFS. It supports multiple compute engines and provides insert, update, and delete (IUD) interfaces and streaming primitives, including upsert and incremental pull, over datasets on HDFS.

NOTE

To use Hudi, ensure that the Spark2x service has been installed in the MRS cluster.

Figure 1-61 Basic architecture of Hudi



Feature

- The ACID transaction capability supports real-time data import to the lake and batch data import to the data lake.
- Multiple view capabilities (read-optimized view/incremental view/real-time view) enable quick data analysis.
- Multi-version concurrency control (MVCC) design supports data version backtracking.
- Automatic management of file sizes and layouts optimizes query performance and provides quasi-real-time data for queries.

- Concurrent read and write are supported. Data can be read when being written based on snapshot isolation.
- Bootstrapping is supported to convert existing tables into Hudi datasets.

Key Technologies and Advantages

- Pluggable index mechanism: Hudi provides multiple index mechanisms to quickly update and delete massive data.
- Ecosystem support: Hudi supports multiple data engines, including Hive, Spark, HetuEngine, and Flink.

Two Types of Tables Supported by Hudi

- Copy On Write

  Copy-on-write tables are also called COW tables. Parquet files are used to store data, and internal update operations need to be performed by rewriting the original Parquet files.

  - Advantage: It is efficient because only one data file in the corresponding partition needs to be read.
  - Disadvantage: During data write, a previous copy needs to be copied and then a new data file is generated based on the previous copy. This process is time-consuming. Therefore, the data read by the read request lags behind.

- Merge On Read

  Merge-on-read tables are also called MOR tables. The combination of columnar-based Parquet and row-based format Avro is used to store data. Parquet files are used to store base data, and Avro files (also called log files) are used to store incremental data.

  - Advantage: Data is written to the delta log first, and the delta log size is small. Therefore, the write cost is low.
  - Disadvantage: Files need to be compacted periodically. Otherwise, there are a large number of fragment files. The read performance is poor because delta logs and old data files need to be merged.

Hudi Supporting Three Types Of Views for Read Capabilities in Different Scenarios

- Snapshot View

  Provides the latest snapshot data of the current Hudi table. That is, once the latest data is written to the Hudi table, the newly written data can be queried through this view.

  Both COW and MOR tables support this view capability.

- Incremental View

  Provides the incremental query capability. The incremental data after a specified commit can be queried. This view can be used to quickly pull incremental data.

  COW tables support this view capability. MOR tables also support this view capability, but the incremental view capability disappears once the compact operation is performed.

- Read Optimized View

  Provides only the data stored in the latest Parquet file.

  This view is different for COW and MOR tables.

  For COW tables, the view capability is the same as the real-time view capability. (COW tables use only Parquet files to store data.)

  For MOR tables, only base files are accessed, and the data in the given file slices since the last compact operation is provided. It can be simply understood that this view provides only the data stored in Parquet files of MOR tables, and the data in log files is ignored. The data provided by this view may not be the latest. However, once the compact operation is performed on MOR tables, the incremental log data is merged into the base data. In this case, this view has the same capability as the real-time view.

1.3.9 Hue

1.3.9.1 Hue Basic Principles

Hue is a group of web applications that interact with MRS big data components. It helps you browse HDFS, perform Hive query, and start MapReduce jobs. Hue bears applications that interact with all MRS big data components.

Hue provides the file browser and query editor functions:

- File browser allows you to directly browse and operate different HDFS directories on the GUI.
- Query editor can write simple SQL statements to query data stored on Hadoop, for example, HDFS, HBase, and Hive. With the query editor, you can easily create, manage, and execute SQL statements and download the execution results as an Excel file.

On the WebUI provided by Hue, you can perform the following operations on the components:

- HDFS:
  - View, create, manage, rename, move, and delete files or directories.
  - File upload and download
  - Search for files, directories, file owners, and user groups; change the owners and permissions of the files and directories.
  - Manually configure HDFS directory storage policies and dynamic storage policies.
- Hive:
  - Edit and execute SQL/HQL statements. Save, copy, and edit the SQL/HQL template. Explain SQL/HQL statements. Save the SQL/HQL statement and query it.
  - Database presentation and data table presentation
  - Supporting different types of Hadoop storage
  - Use MetaStore to add, delete, modify, and query databases, tables, and views.

NOTE

If Internet Explorer is used to access the Hue page to execute HiveSQL statements, the execution fails, because the browser has functional problems. You are advised to use a compatible browser, for example, Google Chrome.

- Impala:
  - Edit and execute SQL/HQL statements. Save, copy, and edit the SQL/HQL template. Explain SQL/HQL statements. Save the SQL/HQL statement and query it.
  - Database presentation and data table presentation
  - Supporting different types of Hadoop storage
  - Use MetaStore to add, delete, modify, and query databases, tables, and views.

  NOTE

  If Internet Explorer is used to access the Hue page to execute HiveSQL statements, the execution fails, because the browser has functional problems. You are advised to use a compatible browser, for example, Google Chrome.
- MapReduce: Check MapReduce tasks that are being executed or have been finished in the clusters, including their status, start and end time, and run logs.
- Oozie: Hue provides the Oozie job manager function, in this case, you can use Oozie in GUI mode.
- ZooKeeper: Hue provides the ZooKeeper browser function for you to use ZooKeeper in GUI mode.

For details about Hue, visit **https://gethue.com/**.

Architecture

Hue, adopting the MTV (Model-Template-View) design, is a web application program running on Django Python. (Django Python is a web application framework that uses open source codes.)

Hue consists of Supervisor Process and WebServer. Supervisor Process is the core Hue process that manages application processes. Supervisor Process and WebServer interact with applications on WebServer through Thrift/REST APIs, as shown in **Figure 1-62**.

Figure 1-62 Hue architecture

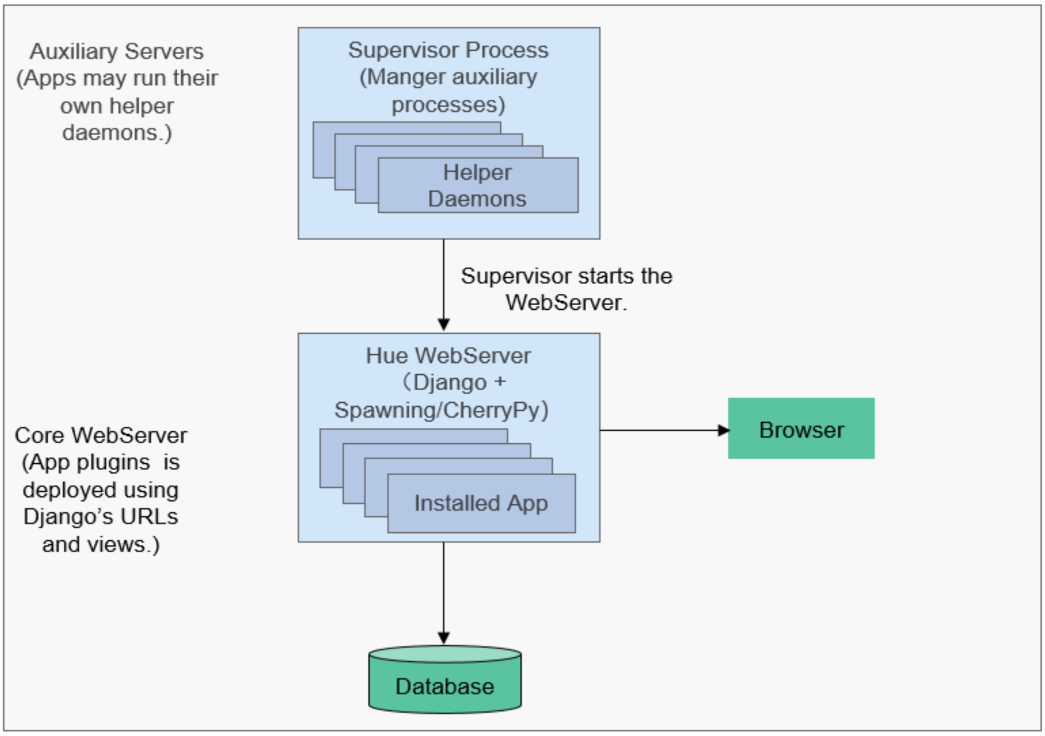

Auxiliary Servers (Apps may run their own helper daemons.)
Supervisor Process (Manger auxiliary processes)
Helper Daemons
Supervisor starts the WebServer.
Hue WebServer（Django + Spawning/CherryPy)
Browser
Core WebServer (App plugins is deployed using Django's URLs and views.)
Installed App
Database

Table 1-11 describes the components shown in **Figure 1-62**.

Table 1-11 Architecture description

Connection Name	Description
Supervisor Process	Manages processes of WebServer applications, such as starting, stopping, and monitoring the processes.
Hue WebServer	Provides the following functions through the Django Python web framework: ● Deploys applications. ● Provides the GUI. ● Connects to databases to store persistent data of applications.

1.3.9.2 Relationship Between Hue and Other Components

Relationship Between Hue and Hadoop Clusters

Figure 1-63 shows how Hue interacts with Hadoop clusters.

Figure 1-63 Hue and Hadoop clusters

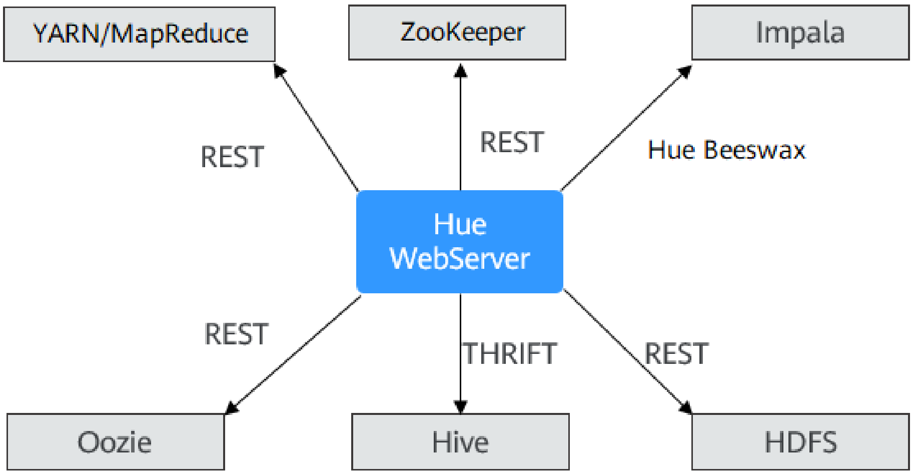


Table 1-12 Relationship Between Hue and Other Components

Connection Name	Description
HDFS	HDFS provides REST APIs to interact with Hue to query and operate HDFS files. Hue packages a user request into interface data, sends the request to HDFS through REST APIs, and displays execution results on the web UI.
Hive	Hive provides Thrift interfaces to interact with Hue, execute Hive SQL statements, and query table metadata. If you edit HQL statements on the Hue web UI, then, Hue submits the HQL statements to the Hive server through the Thrift APIs and displays execution results on the web UI.
YARN/ MapReduce	MapReduce provides REST APIs to interact with Hue and query YARN job information. If you go to the Hue web UI, enter the filter parameters, the UI sends the parameters to the background, and Hue invokes the REST APIs provided by MapReduce (MR1/MR2-YARN) to obtain information such as the status of the task running, the start/end time, the run log, and more.
Oozie	Oozie provides REST APIs to interact with Hue, create workflows, coordinators, and bundles, and manage and monitor tasks. A graphical workflow, coordinator, and bundle editor are provided on the Hue web UI. Hue invokes the REST APIs of Oozie to create, modify, delete, submit, and monitor workflows, coordinators, and bundles.

Connection Name	Description
ZooKeeper	ZooKeeper provides REST APIs to interact with Hue and query ZooKeeper node information. ZooKeeper node information is displayed in the Hue web UI. Hue invokes the REST APIs of ZooKeeper to obtain the node information.
Impala	Impala provides Hue Beeswax APIs to interact with Hue, execute Hive SQL statements, and query table metadata. If you edit HQL statements on the Hue web UI, then, Hue submits the HQL statements to the Hive server through the Hue Beeswax APIs and displays execution results on the web UI.

1.3.9.3 Hue Enhanced Open Source Features

Hue Enhanced Open Source Features

- Storage policy: The number of HDFS file copies varies depending on the storage media. This feature allows you to manually set an HDFS directory storage policy or can automatically adjust the file storage policy, modify the number of file copies, move the file directory, and delete files based on the latest access time and modification time of HDFS files to fully utilize storage capacity and improve storage performance.
- MR engine: You can use the MapReduce engine to execute Hive SQL statements.
- Reliability enhancement: Hue is deployed in active/standby mode. When interconnecting with HDFS, Oozie, Hive, and YARN, Hue can work in failover or load balancing mode.

1.3.10 Impala

Impala

Impala provides fast, interactive SQL queries directly on your Apache Hadoop data stored in HDFS, HBase, or the Object Storage Service (OBS). In addition to using the same unified storage platform, Impala also uses the same metadata, SQL syntax (Hive SQL), ODBC driver, and user interface (Impala query UI in Hue) as Apache Hive. This provides a familiar and unified platform for real-time or batch-oriented queries. Impala is an addition to tools available for querying big data. Impala does not replace the batch processing frameworks built on MapReduce such as Hive. Hive and other frameworks built on MapReduce are best suited for long running batch jobs.

Impala provides the following features:

- Most common SQL-92 features of Hive Query Language (HiveQL) including SELECT, JOIN, and aggregate functions

- HDFS, HBase, and OBS storage, including:
  - HDFS file formats: delimited text files, Parquet, Avro, SequenceFile, and RCFile
  - Compression codecs: Snappy, GZIP, Deflate, BZIP
- Common data access interfaces including:
  - JDBC driver
  - ODBC driver
  - Hue Beeswax and the Impala query UI
- **impala-shell** command line interface
- Kerberos authentication

Impala applies to offline analysis (such as log and cluster status analysis) of real-time data queries, large-scale data mining (such as user behavior analysis, interest region analysis, and region display), and other scenarios.

For details about Impala, visit **https://impala.apache.org/impala-docs.html**.

Impala consists of three roles: Impala Daemon (Impalad), Impala StateStore, and Impala Catalog Service.

Impala Daemon

The core Impala component is the Impala daemon, physically represented by the **impalad** process.

A few of the key functions that an Impala daemon performs are:

- Runs on all data nodes.
- Reads and writes to data files.
- Accepts queries transmitted from the **impala-shell** command, Hue, JDBC, or ODBC.
- Parallelizes the queries and transmits intermediate query results back to the central coordinator.
- Invokes a node to return the query results to the client.

The Impala daemons are in constant communication with StateStore, to confirm which daemons are healthy and can accept new work.

Impala StateStore

The Impala component known as the StateStore checks on the health of all Impala daemons in a cluster, and continuously relays its findings to each of those daemons. It is physically represented by a daemon process named **statestored**. You only need such a process on one host in a cluster. If an Impala daemon goes offline due to hardware failure, network error, software issue, or other reason, the StateStore informs all the other Impala daemons so that future queries can avoid making requests to the unreachable Impala daemon.

Impala Catalog Service

The Impala component known as the Catalog Service relays the metadata changes from Impala SQL statements to all the Impala daemons in a cluster. It is physically represented by a daemon process named **catalogd**. When you create a table, load data, and so on through Hive, you do need to issue REFRESH or INVALIDATE METADATA on an Impala daemon before executing a query there. The catalog service avoids the need to issue REFRESH and INVALIDATE METADATA statements when the metadata changes are performed by statements issued through Impala.

1.3.11 Kafka

1.3.11.1 Kafka Basic Principles

Kafka is an open source, distributed, partitioned, and replicated commit log service. Kafka is publish-subscribe messaging, rethought as a distributed commit log. It provides features similar to Java Message Service (JMS) but another design. It features message endurance, high throughput, distributed methods, multi-client support, and real time. It applies to both online and offline message consumption, such as regular message collection, website activeness tracking, aggregation of statistical system operation data (monitoring data), and log collection. These scenarios engage large amounts of data collection for Internet services.

Kafka Structure

Producers publish data to topics, and consumers subscribe to the topics and consume messages. A broker is a server in a Kafka cluster. For each topic, the Kafka cluster maintains partitions for scalability, parallelism, and fault tolerance. Each partition is an ordered, immutable sequence of messages that is continually appended to - a commit log. Each message in a partition is assigned a sequential ID, which is called offset.

Figure 1-64 Kafka architecture

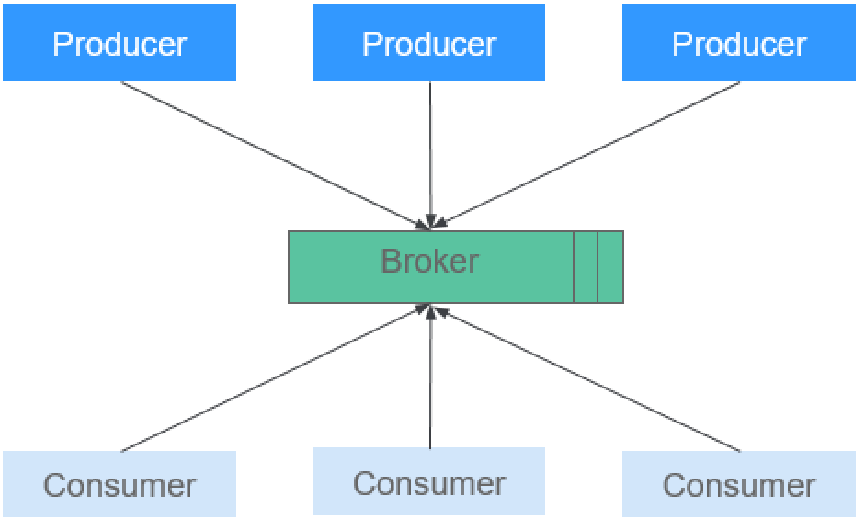
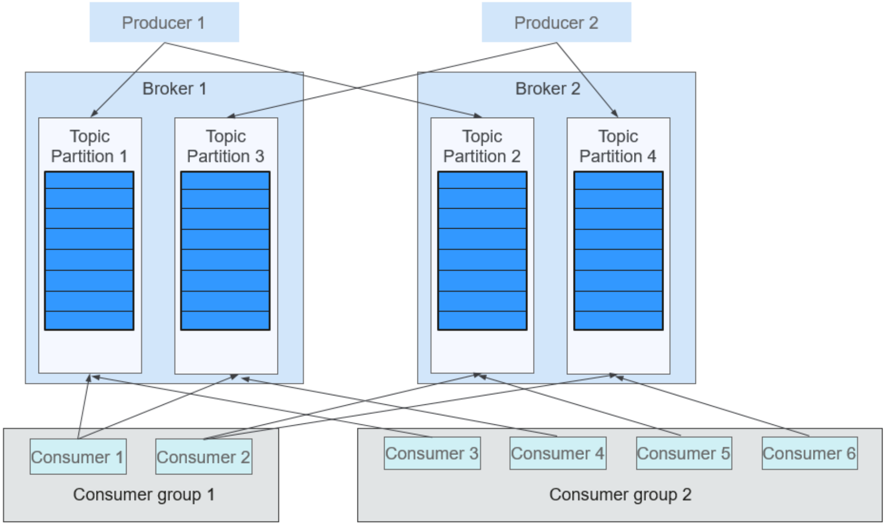


Table 1-13 Kafka architecture description

Name	Description
Broker	A broker is a server in a Kafka cluster.
Topic	A topic is a category or feed name to which messages are published. A topic can be divided into multiple partitions, which can act as a parallel unit.
Partition	A partition is an ordered, immutable sequence of messages that is continually appended to - a commit log. The messages in the partitions are each assigned a sequential ID number called the offset that uniquely identifies each message within the partition.
Producer	Producers publish messages to a Kafka topic.
Consumer	Consumers subscribe to topics and process the feed of published messages.

Figure 1-65 shows the relationships between modules.

Figure 1-65 Relationships between Kafka modules



Consumers label themselves with a consumer group name, and each message published to a topic is delivered to one consumer instance within each subscribing consumer group. If all the consumer instances belong to the same consumer group, loads are evenly distributed among the consumers. As shown in the preceding figure, Consumer1 and Consumer2 work in load-sharing mode; Consumer3, Consumer4, Consumer5, and Consumer6 work in load-sharing mode. If all the consumer instances belong to different consumer groups, messages are broadcast to all consumers. As shown in the preceding figure, the messages in Topic 1 are broadcast to all consumers in Consumer Group1 and Consumer Group2.

For details about Kafka architecture and principles, see **https://kafka.apache.org/24/documentation.html**.

Principle

- **Message Reliability**

  When a Kafka broker receives a message, it stores the message on a disk persistently. Each partition of a topic has multiple replicas stored on different broker nodes. If one node is faulty, the replicas on other nodes can be used.
- **High Throughput**

  Kafka provides high throughput in the following ways:
  - Messages are written into disks instead of being cached in the memory, fully utilizing the sequential read and write performance of disks.
  - The use of zero-copy eliminates I/O operations.
  - Data is sent in batches, improving network utilization.
  - Each topic is divided in to multiple partitions, which increases concurrent processing. Concurrent read and write operations can be performed between multiple producers and consumers. Producers send messages to specified partitions based on the algorithm used.
- **Message Subscribe-Notify Mechanism**

  Consumers subscribe to interested topics and consume data in pull mode. Consumers can choose the consumption mode, such as batch consumption, repeated consumption, and consumption from the end, and control the message pulling speed based on actual situation. Consumers need to maintain the consumption records by themselves.
- **Scalability**

  When broker nodes are added to expand the Kafka cluster capacity, the newly added brokers register with ZooKeeper. After the registration is successful, procedures and consumers can sense the change in a timely manner and make related adjustment.

Open Source Features

- Reliability

  Message processing methods such as **At-Least Once**, **At-Most Once**, and **Exactly Once** are provided. The message processing status is maintained by consumers. Kafka needs to work with the application layer to implement **Exactly Once**.
- High throughput

  High throughput is provided for message publishing and subscription.
- Persistence

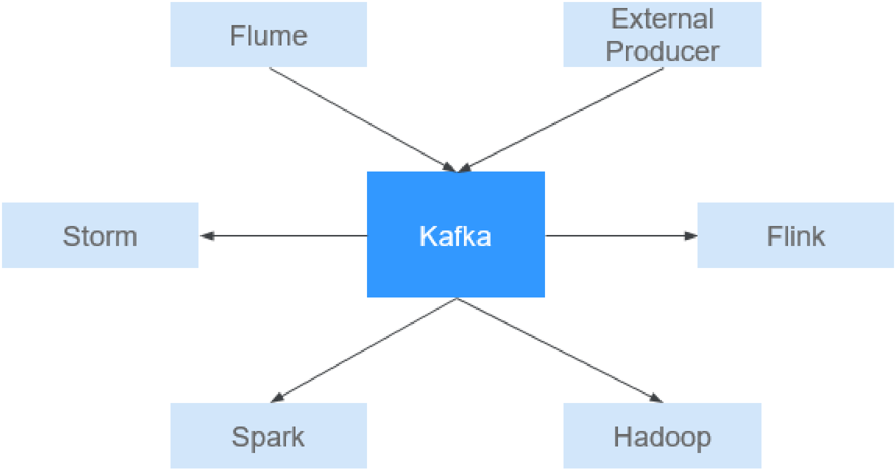
  Messages are stored on disks and can be used for batch consumption and real-time application programs. Data persistence and replication prevent data loss.
- Distribution

  A distributed system is easy to be expanded externally. All producers, brokers, and consumers support the deployment of multiple distributed clusters. Systems can be scaled without stopping the running of software or shutting down the machines.

1.3.11.2 Relationship Between Kafka and Other Components

As a message publishing and subscription system, Kafka provides high-speed data transmission methods for data transmission between different subsystems of the FusionInsight platform. It can receive external messages in a real-time manner and provides the messages to the online and offline services for processing. The following figure shows the relationship between Kafka and other components.

Figure 1-66 Relationship with Other Components



1.3.11.3 Kafka Enhanced Open Source Features

Kafka Enhanced Open Source Features

- Monitors the following topic-level metrics:
  - Topic Input Traffic
  - Topic Output Traffic
  - Topic Rejected Traffic
  - Number of Failed Fetch Requests Per Second
  - Number of Failed Produce Requests Per Second
  - Number of Topic Input Messages Per Second
  - Number of Fetch Requests Per Second
  - Number of Produce Requests Per Second
- Queries the mapping between broker IDs and node IP addresses. On Linux clients, **kafka-broker-info.sh** can be used to query the mapping between broker IDs and node IP addresses.

1.3.12 KafkaManager

KafkaManager is a tool for managing Apache Kafka and provides GUI-based metric monitoring and management of Kafka clusters.

KafkaManager supports the following operations:

- Manage multiple Kafka clusters.
- Easy inspection of cluster states (topics, consumers, offsets, partitions, replicas, and nodes)
- Run preferred replica election.
- Generate partition assignments with option to select brokers to use.
- Run reassignment of partition (based on generated assignments).
- Create a topic with optional topic configurations (Multiple Kafka cluster versions are supported).
- Delete a topic (only supported on 0.8.2+ and **delete.topic.enable=true** is set in broker configuration).
- Batch generate partition assignments for multiple topics with option to select brokers to use.
- Batch run reassignment of partitions for multiple topics.
- Add partitions to an existing topic.
- Update configurations for an existing topic.
- Optionally enable JMX polling for broker-level and topic-level metrics.
- Optionally filter out consumers that do not have ids/ owner / & offsets/ directories in ZooKeeper.

1.3.13 KrbServer and LdapServer

1.3.13.1 KrbServer and LdapServer Principles

Overview

To manage the access control permissions on data and resources in a cluster, it is recommended that the cluster be installed in security mode. In security mode, a client application must be authenticated and a secure session must be established before the application accesses any resource in the cluster. MRS uses KrbServer to provide Kerberos authentication for all components, implementing a reliable authentication mechanism.

LdapServer supports Lightweight Directory Access Protocol (LDAP) and provides the capability of storing user and user group data for Kerberos authentication.

Architecture

The security authentication function for user login depends on Kerberos and LDAP.

Figure 1-67 Security authentication architecture

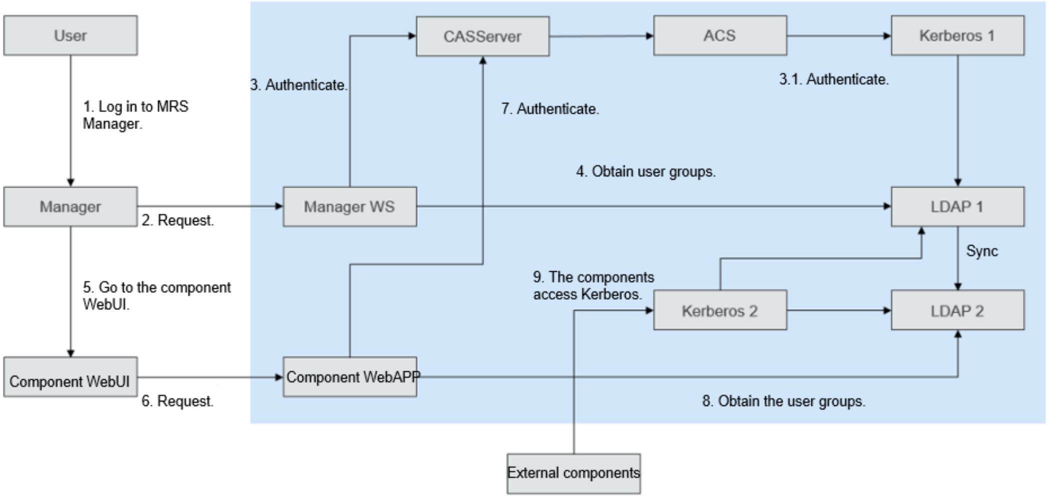


Figure 1-67 includes three scenarios:

- Logging in to the MRS Manager Web UI

  The authentication architecture includes steps 1, 2, 3, and 4.
- Logging in to a component web UI

  The authentication architecture includes steps 5, 6, 7, and 8.
- Accessing between components

  The authentication architecture includes step 9.

Table 1-14 Key modules

Connection Name	Description
Manager	Cluster Manager
Manager WS	WebBrowser
Kerberos1	KrbServer (management plane) service deployed in MRS Manager, that is, OMS Kerberos
Kerberos2	KrbServer (service plane) service deployed in the cluster
LDAP1	LdapServer (management plane) service deployed in MRS Manager, that is, OMS LDAP
LDAP2	LdapServer (service plane) service deployed in the cluster

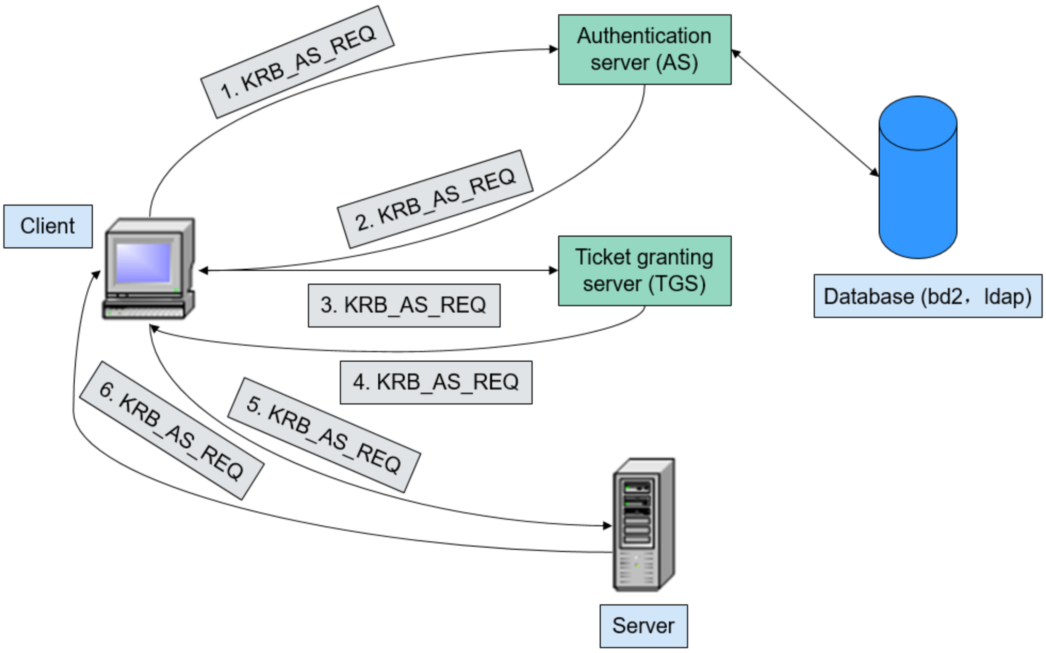
Data operation mode of Kerberos1 in LDAP: The active and standby instances of LDAP1 and the two standby instances of LDAP2 can be accessed in load balancing mode. Data write operations can be performed only in the active LDAP1 instance. Data read operations can be performed in LDAP1 or LDAP2.

Data operation mode of Kerberos2 in LDAP: Data read operations can be performed in LDAP1 and LDAP2. Data write operations can be performed only in the active LDAP1 instance.

Principle

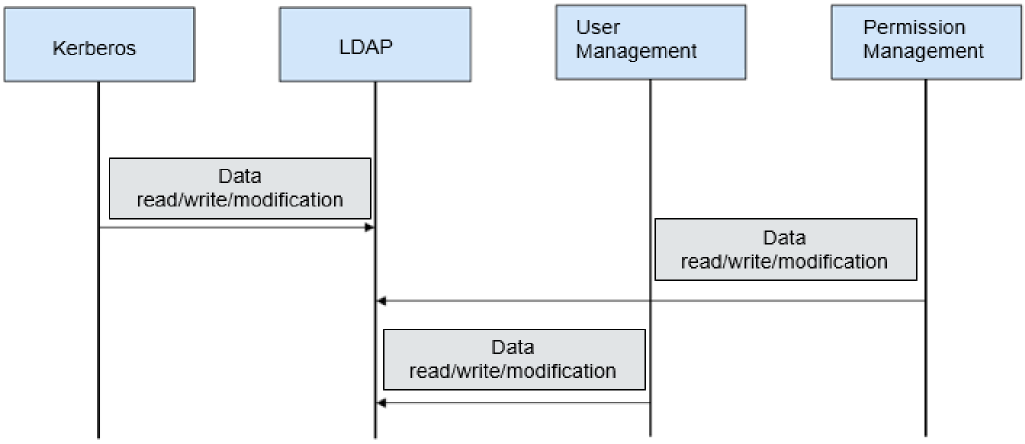
Kerberos authentication

Figure 1-68 Authentication process



LDAP data read and write

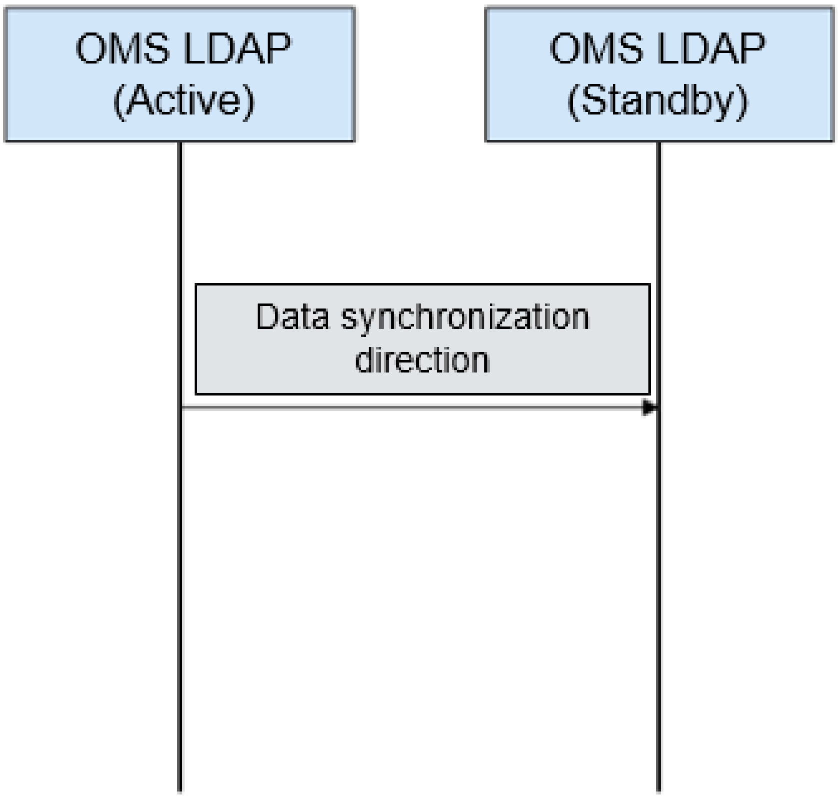
Figure 1-69 Data modification process


Kerberos
LDAP
User Management
Permission Management
Data read/write/modification
Data read/write/modification
Data read/write/modification

LDAP data synchronization

- OMS LDAP data synchronization before cluster installation

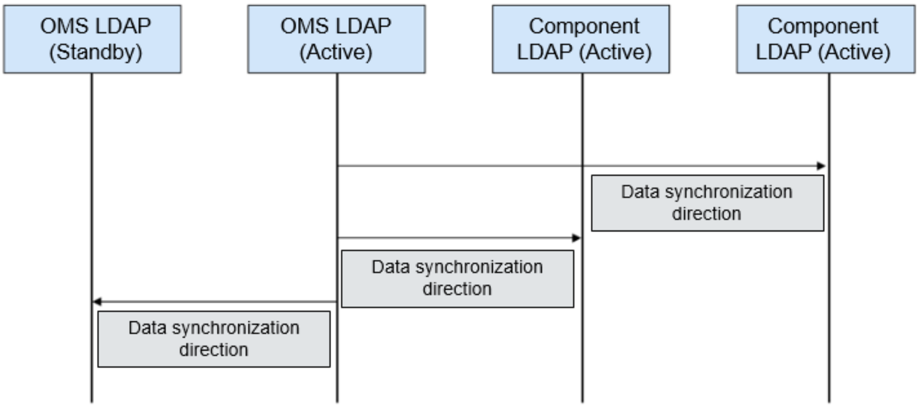
Figure 1-70 OMS LDAP data synchronization



Data synchronization direction before cluster installation: Data is synchronized from the active OMS LDAP to the standby OMS LDAP.

- LDAP data synchronization after cluster installation

Figure 1-71 LDAP data synchronization


OMS LDAP (Standby)
OMS LDAP (Active)
Component LDAP (Active)
Component LDAP (Active)
Data synchronization direction
Data synchronization direction
Data synchronization direction

Data synchronization direction after cluster installation: Data is synchronized from the active OMS LDAP to the standby OMS LDAP, standby component LDAP, and standby component LDAP.

1.3.13.2 KrbServer and LdapServer Enhanced Open Source Features

Enhanced open-source features of KrbServer and LdapServer: intra-cluster service authentication

In an MRS cluster that uses the security mode, mutual access between services is implemented based on the Kerberos security architecture. When a service (such as HDFS) in the cluster is to be started, the corresponding sessionkey (keytab, used for identity authentication of the application) is obtained from Kerberos. If another service (such as YARN) needs to access HDFS and add, delete, modify, or query data in HDFS, the corresponding TGT and ST must be obtained for secure access.

Enhanced Open-Source Features of KrbServer and LdapServer: Application Development Authentication

MRS components provide application development interfaces for customers or upper-layer service product clusters. During application development, a cluster in security mode provides specified application development authentication interfaces to implement application security authentication and access. For example, the UserGroupInformation class provided by the hadoop-common API provides multiple security authentication APIs.

- **setConfiguration()** is used to obtain related configuration and set parameters such as global variables.
- **loginUserFromKeytab():** is used to obtain TGT interfaces.

Enhanced Open-Source Features of KrbServer and LdapServer: Cross-System Mutual Trust

MRS provides the mutual trust function between two Managers to implement data read and write operations between systems.

1.3.14 Kudu

Kudu is a columnar storage manager developed for the Apache Hadoop platform. Kudu shares the common technical properties of Hadoop ecosystem applications: it runs on commodity hardware, is horizontally scalable, and supports highly available operation.

Kudu's design has the following benefits:

- Fast processing of OLAP workloads
- Integration with MapReduce, Spark and other Hadoop ecosystem components
- Tight integration with Apache Impala, making it a good, mutable alternative to using HDFS with Apache Parquet
- Strong but flexible consistency model, allowing you to choose consistency requirements on a per-request basis, including the option for strict-serializable consistency
- Strong performance for running sequential and random workloads simultaneously

- Easy to manage
- High availability Tablet Servers and Masters use the Raft Consensus Algorithm, which ensures that as long as more than half the total number of replicas is available, the tablet is available for reads and writes. For example, if 2 out of 3 replicas or 3 out of 5 replicas are available, the tablet is available. Reads can be serviced by read-only follower tablets, even in the event of a leader tablet failure.
- Structured data model

By combining all of these properties, Kudu targets support for families of applications that are difficult or impossible to implement on current generation Hadoop storage technologies.

A few examples of applications for which Kudu is a great solution are:

- Reporting applications where newly-arrived data needs to be immediately available for end users
- Time-series applications that must simultaneously support queries across large amounts of historic data and granular queries about an individual entity that must return very quickly
- Applications that use predictive models to make real-time decisions with periodic refreshes of the predictive model based on all historic data

1.3.15 Loader

1.3.15.1 Loader Basic Principles

Loader is developed based on the open source Sqoop component. It is used to exchange data and files between MRS and relational databases and file systems. Loader can import data from relational databases or file servers to the HDFS and HBase components, or export data from HDFS and HBase to relational databases or file servers.

A Loader model consists of Loader Client and Loader Server, as shown in **Figure 1-72**.

Figure 1-72 Loader model

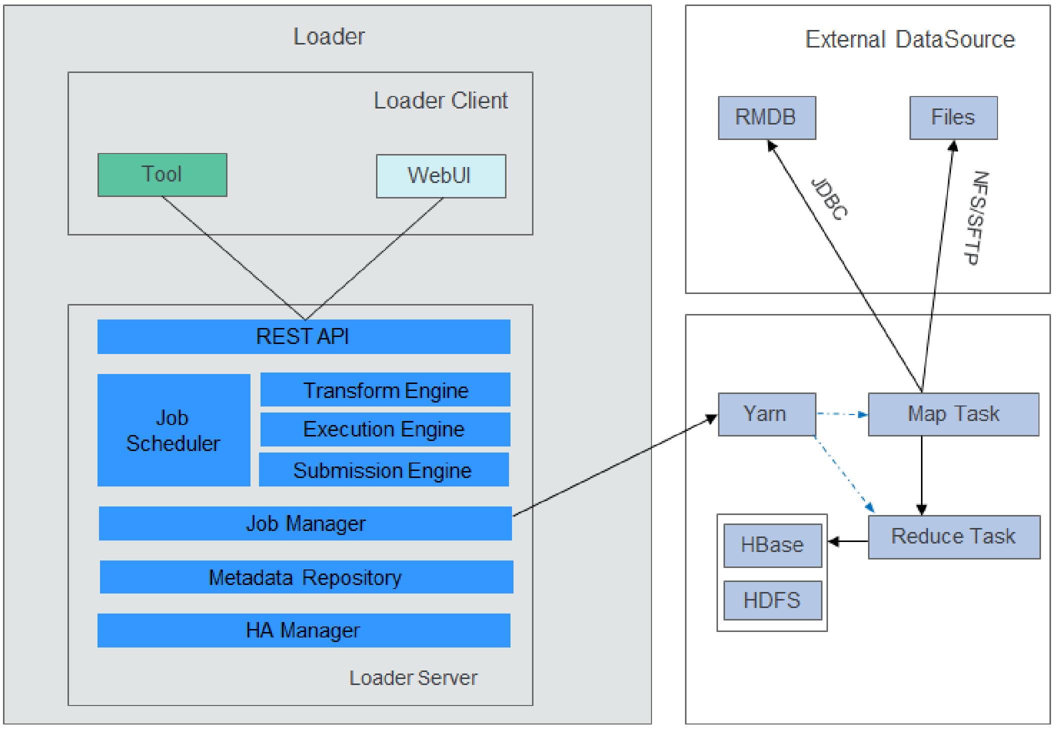


Table 1-15 describes the functions of each module shown in the preceding figure.

Table 1-15 Components of the Loader model

Module	Description
Loader Client	Loader client. It provides two interfaces: web UI and CLI.
Loader Server	Loader server. It processes operation requests sent from the client, manages connectors and metadata, submits MapReduce jobs, and monitors MapReduce job status.
REST API	It provides a Representational State Transfer (RESTful) APIs (HTTP + JSON) to process the operation requests sent from the client.
Job Scheduler	Simple job scheduler. It periodically executes Loader jobs.
Transform Engine	Data transformation engine. It supports field combination, string cutting, and string reverse.
Execution Engine	Loader job execution engine. It executes Loader jobs in MapReduce manner.
Submission Engine	Loader job submission engine. It submits Loader jobs to MapReduce.
Job Manager	It manages Loader jobs, including creating, querying, updating, deleting, activating, deactivating, starting, and stopping jobs.

Module	Description
Metadata Repository	Metadata repository. It stores and manages data about Loader connectors, transformation procedures, and jobs.
HA Manager	It manages the active/standby status of Loader Server processes. The Loader Server has two nodes that are deployed in active/standby mode.

Loader imports or exports jobs in parallel using MapReduce jobs. Some job import or export may involve only the Map operations, while some may involve both Map and Reduce operations.

Loader implements fault tolerance using MapReduce. Jobs can be rescheduled upon a job execution failure.

- **Importing data to HBase**

  When the Map operation is performed for MapReduce jobs, Loader obtains data from an external data source.

  When a Reduce operation is performed for a MapReduce job, Loader enables the same number of Reduce tasks based on the number of Regions. The Reduce tasks receive data from Map tasks, generate HFiles by Region, and store the HFiles in a temporary directory of HDFS.

  When a MapReduce job is submitted, Loader migrates HFiles from the temporary directory to the HBase directory.

- **Importing Data to HDFS**

  When a Map operation is performed for a MapReduce job, Loader obtains data from an external data source and exports the data to a temporary directory (named *export directory*-**ldtmp**).

  When a MapReduce job is submitted, Loader migrates data from the temporary directory to the output directory.

- **Exporting data to a relational database**

  When a Map operation is performed for a MapReduce job, Loader obtains data from HDFS or HBase and inserts the data to a temporary table (Staging Table) through the Java DataBase Connectivity (JDBC) API.

  When a MapReduce job is submitted, Loader migrates data from the temporary table to a formal table.

- **Exporting data to a file system**

  When a Map operation is performed for a MapReduce job, Loader obtains data from HDFS or HBase and writes the data to a temporary directory of the file server.

  When a MapReduce job is submitted, Loader migrates data from the temporary directory to a formal directory.

For details about the Loader architecture and principles, see **https://sqoop.apache.org/docs/1.99.3/index.html**.

1.3.15.2 Relationship Between Loader and Other Components

The components that interact with Loader include HDFS, HBase, MapReduce, and ZooKeeper. Loader works as a client to use certain functions of these components, such as storing data to HDFS and HBase and reading data from HDFS and HBase tables. In addition, Loader functions as an MapReduce client to import or export data.

1.3.15.3 Loader Enhanced Open Source Features

Loader Enhanced Open-Source Feature: Data Import and Export

Loader is developed based on Sqoop. In addition to the Sqoop functions, Loader has the following enhanced features:

- Provides data conversion functions.
- Supports GUI-based configuration conversion.
- Imports data from an SFTP/FTP server to HDFS/OBS.
- Imports data from an SFTP/FTP server to an HBase table.
- Imports data from an SFTP/FTP server to a Phoenix table.
- Imports data from an SFTP/FTP server to a Hive table.
- Exports data from HDFS/OBS to an SFTP/FTP server.
- Exports data from an HBase table to an SFTP/FTP server.
- Exports data from a Phoenix table to an SFTP/FTP server.
- Imports data from a relational database to an HBase table.
- Imports data from a relational database to a Phoenix table.
- Imports data from a relational database to a Hive table.
- Exports data from an HBase table to a relational database.
- Exports data from a Phoenix table to a relational database.
- Imports data from an Oracle partitioned table to HDFS/OBS.
- Imports data from an Oracle partitioned table to an HBase table.
- Imports data from an Oracle partitioned table to a Phoenix table.
- Imports data from an Oracle partitioned table to a Hive table.
- Exports data from HDFS/OBS to an Oracle partitioned table.
- Exports data from HBase to an Oracle partitioned table.
- Exports data from a Phoenix table to an Oracle partitioned table.
- Imports data from HDFS to an HBase table, a Phoenix table, and a Hive table in the same cluster.
- Exports data from an HBase table and a Phoenix table to HDFS/OBS in the same cluster.
- Imports data to an HBase table and a Phoenix table by using **bulkload** or **put list**.
- Imports all types of files from an SFTP/FTP server to HDFS. The open source component Sqoop can import only text files.
- Exports all types of files from HDFS/OBS to an SFTP server. The open source component Sqoop can export only text files and SequenceFile files.

- Supports file coding format conversion during file import and export. The supported coding formats include all formats supported by Java Development Kit (JDK).
- Retains the original directory structure and file names during file import and export.
- Supports file combination during file import and export. For example, if a large number of files are to be imported, these files can be combined into *n* files (*n* can be configured).
- Supports file filtering during file import and export. The filtering rules support wildcards and regular expressions.
- Supports batch import and export of ETL tasks.
- Supports query by page and key word and group management of ETL tasks.
- Provides floating IP addresses for external components.

1.3.16 Manager

1.3.16.1 Manager Basic Principles

Overview

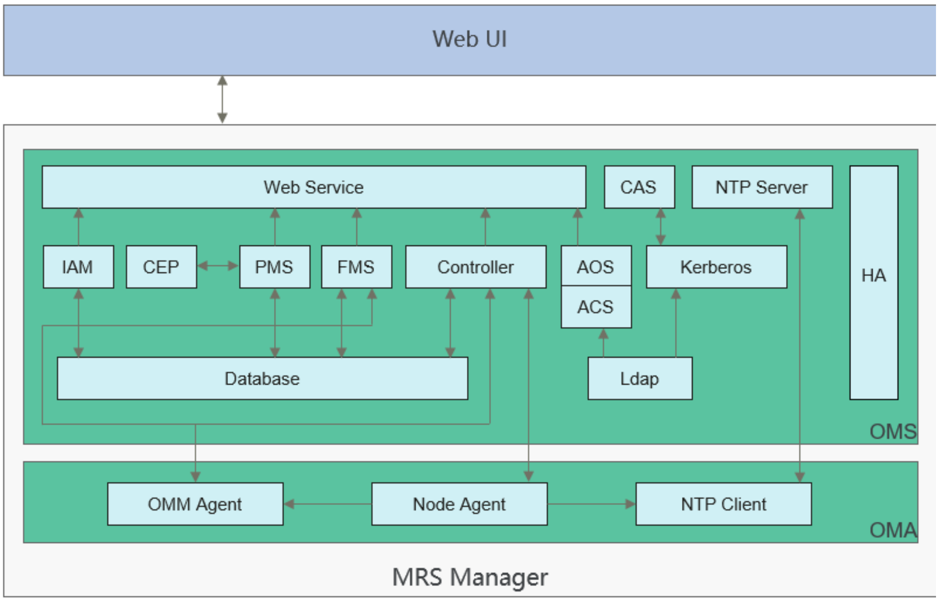
Manager is the O&M management system of MRS and provides unified cluster management capabilities for services deployed in clusters.

Manager provides functions such as performance monitoring, alarms, user management, permission management, auditing, service management, health check, and log collection.

Architecture

Figure 1-73 shows the overall logical architecture of FusionInsight Manager.

Figure 1-73 Manager logical architecture



Manager consists of OMS and OMA.

- OMS: serves as management node in the O&M system. There are two OMS nodes deployed in active/standby mode.
- OMA: managed node in the O&M system. Generally, there are multiple OMA nodes.

Figure 1-73 describes the modules shown in **Table 1-16**.

Table 1-16 Service module description

Module	Description
Web Service	A web service deployed under Tomcat, providing HTTPS API of Manager. It is used to access Manager through the web browser. In addition, it provides the northbound access capability based on the Syslog and SNMP protocols.
OMS	Management node of the O&M system. Generally, there are two OMS nodes that work in active/standby mode.
OMA	Managed node in the O&M system. Generally, there are multiple OMA nodes.
Controller	The control center of Manager. It can converge information from all nodes in the cluster and display it to administrators, as well as receive from administrators, and synchronize information to all nodes in the cluster according to the operation instruction range. Control process of Manager. It implements various management actions: 1. The web service delivers various management actions (such as installation, service startup and stop, and configuration modification) to Controller. 2. Controller decomposes the command and delivers the action to each Node Agent, for example, starting a service involves multiple roles and instances. 3. Controller is responsible for monitoring the implementation of each action.
Node Agent	Node Agent exists on each cluster node and is an enabler of Manager on a single node. ● Node Agent represents all the components deployed on the node to interact with Controller, implementing convergence from multiple nodes of a cluster to a single node. ● Node Agent enables Controller to perform all operations on the components deployed on the node. It allows Controller functions to be implemented. Node Agent sends heartbeat messages to Controller at an interval of 3 seconds. The interval cannot be configured.
IAM	Records audit logs. Each non-query operation on the Manager UI has a related audit log.

Module	Description
PMS	The performance monitoring module. It collects the performance monitoring data on each OMA and provides the query function.
CEP	Convergence function module. For example, the used disk space of all OMAs is collected as a performance indicator.
FMS	Alarm module. It collects and queries alarms on each OMA.
OMM Agent	Agent for performance monitoring and alarm reporting on the OMA. It collects performance monitoring data and alarm data on Agent Node.
CAS	Unified authentication center. When a user logs in to the web service, CAS authenticates the login. The browser automatically redirects the user to the CAS through URLs.
AOS	Permission management module. It manages the permissions of users and user groups.
ACS	User and user group management module. It manages users and user groups to which users belong.
Kerberos	LDAP is deployed in OMS and a cluster, respectively. ● OMS Kerberos provides the single sign-on (SSO) and authentication between Controller and Node Agent. ● Kerberos in the cluster provides the user security authentication function for components. The service name is **KrbServer**, which contains two role instances: – KerberosServer: is an authentication server that provides security authentication for MRS. – KerberosAdmin: manages processes of Kerberos users.
Ldap	LDAP is deployed in OMS and a cluster, respectively. ● OMS LDAP provides data storage for user authentication. ● The LDAP in the cluster functions as the backup of the OMS LDAP. The service name is **LdapServer** and the role instance is **SlapdServer**.
Database	Manager database used to store logs and alarms.
HA	HA management module that manages the active and standby OMSs.
NTP Server NTP Client	It synchronizes the system clock of each node in the cluster.

1.3.16.2 Manager Key Features

Key Feature: Unified Alarm Monitoring

Manager provides the visualized and convenient alarm monitoring function. Users can quickly obtain key cluster performance indicators, evaluate cluster health status, customize performance indicator display, and convert indicators to alarms. Manager can monitor the running status of all components and report alarms in real time when faults occur. The online help on the GUI allows you to view performance counters and alarm clearance methods to quickly rectify faults.

Key Feature: Unified User Permission Management

Manager provides permission management of components in a unified manner.

Manager introduces the concept of role and uses role-based access control (RBAC) to manage system permissions. It centrally displays and manages scattered permission functions of each component in the system and organizes the permissions of each component in the form of permission sets (roles) to form a unified system permission concept. By doing so, common users cannot obtain internal permission management details, and permissions become easy for administrators to manage, greatly facilitating permission management and improving user experience.

Key Feature: SSO

Single sign-on (SSO) is provided between the Manager web UI and component web UI as well as for integration between MRS and third-party systems.

This function centrally manages and authenticates Manager users and component users. The entire system uses LDAP to manage users and uses Kerberos for authentication. A set of Kerberos and LDAP management mechanisms are used between the OMS and components. SSO (including single sign-on and single sign-out) is implemented through CAS. With SSO, users can easily switch tasks between the Manager web UI, component web UIs, and third-party systems, without switching to another user.

NOTE

- To ensure security, the CAS Server can retain a ticket-granting ticket (TGT) used by a user only for 20 minutes.
- If a user does not perform any operation on the page (including on the Manager web UI and component web UIs) within 20 minutes, the page is automatically locked.

Key Feature: Automatic Health Check and Inspection

Manager provides users with automatic inspection on system running environments and helps users check and audit system running health by one click, ensuring correct system running and lowering system operation and maintenance costs. After viewing inspection results, you can export reports for archiving and fault analysis.

Key Feature: Tenant Management

Manager introduces the multi-tenant concept. The CPU, memory, and disk resources of a cluster can be integrated into a set. The set is called a tenant. A mode involving different tenants is called multi-tenant mode.

Manager provides the multi-tenant function, supports a level-based tenant model and allows tenants to be added and deleted dynamically, achieving resource isolation. As a result, it can dynamically manage and configure the computing resources and the storage resources of tenants.

- The computing resources indicate tenants' Yarn task queue resources. The task queue quota can be modified, and the task queue usage status and statistics can be viewed.
- The storage resources can be stored on HDFS. You can add and delete the HDFS storage directories of tenants, and set the quotas of file quantity and the storage space of the directories.

As a unified tenant management platform of MRS, MRS Manager allows users to create and manage tenants in clusters based on service requirements.

- Roles, computing resources, and storage resources are automatically created when tenants are created. By default, all permissions of the new computing resources and storage resources are allocated to a tenant's roles.
- After you have modified the tenant's computing or storage resources, permissions of the tenant's roles are automatically updated.

Manager also provides the multi-instance function so that users can use the HBase, Hive, or Spark alone in the resource control and service isolation scenario. The multi-instance function is disabled by default and can be manually enabled.

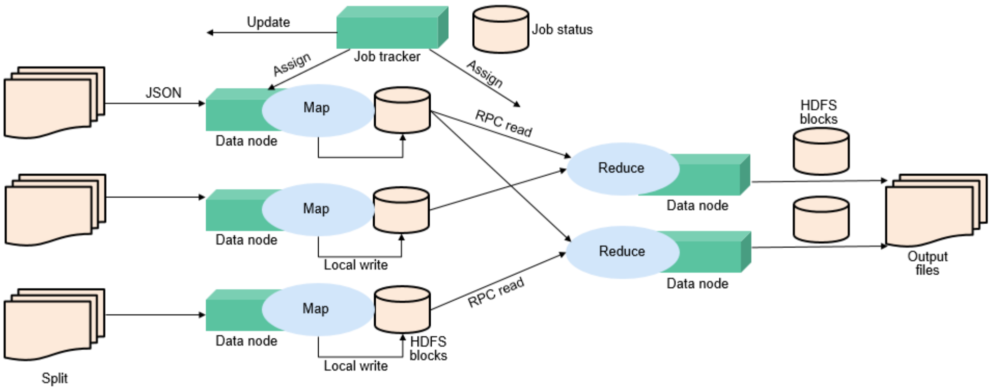
1.3.17 MapReduce

1.3.17.1 MapReduce Basic Principles

MapReduce is the core of Hadoop. As a software architecture proposed by Google, MapReduce is used for parallel computing of large-scale datasets (larger than 1 TB). The concepts "Map" and "Reduce" and their main thoughts are borrowed from functional programming language and also borrowed from the features of vector programming language.

Current software implementation is as follows: Specify a Map function to map a series of key-value pairs into a new series of key-value pairs, and specify a Reduce function to ensure that all values in the mapped key-value pairs share the same key.

Figure 1-74 Distributed batch processing engine



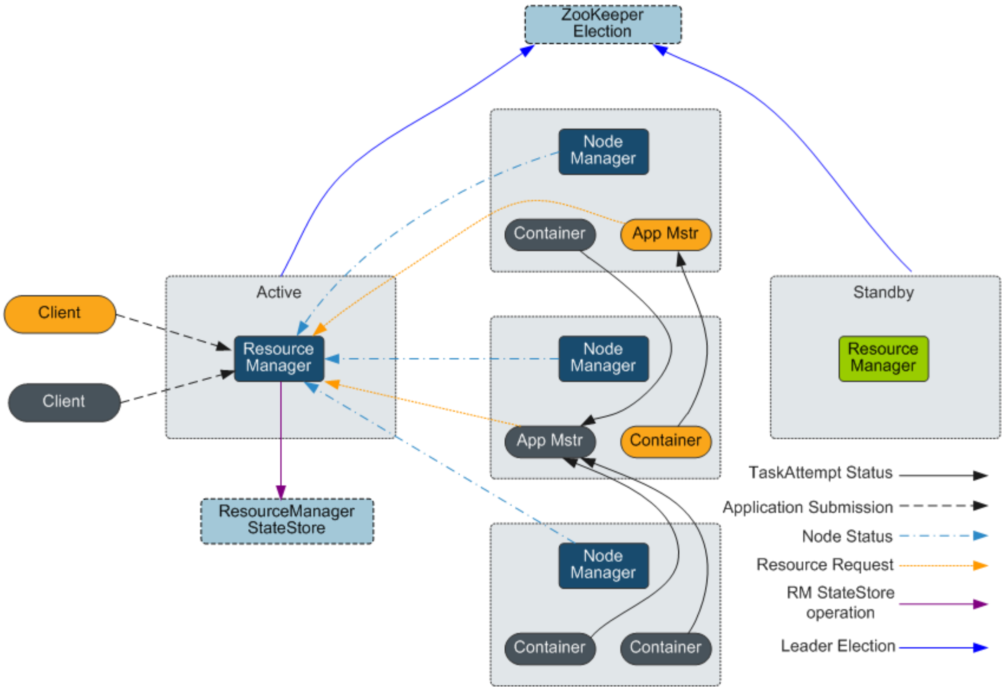
MapReduce is a software framework for processing large datasets in parallel. The root of MapReduce is the Map and Reduce functions in functional programming. The Map function accepts a group of data and transforms it into a key-value pair list. Each element in the input domain corresponds to a key-value pair. The Reduce function accepts the list generated by the Map function, and then shrinks the key-value pair list based on the keys. MapReduce divides a task into multiple parts and allocates them to different devices for processing. In this way, the task can be finished in a distributed environment instead of a single powerful server.

For more information, see **MapReduce Tutorial**.

MapReduce structure

As shown in **Figure 1-75**, MapReduce is integrated into YARN through the Client and ApplicationMaster interfaces of YARN, and uses YARN to apply for computing resources.

Figure 1-75 Basic architecture of Apache YARN and MapReduce



1.3.17.2 Relationship Between MapReduce and Other Components

Relationship Between MapReduce and HDFS

- HDFS features high fault tolerance and high throughput, and can be deployed on low-cost hardware for storing data of applications with massive data sets.
- MapReduce is a programming model used for parallel computation of large data sets (larger than 1 TB). Data computed by MapReduce comes from multiple data sources, such as Local FileSystem, HDFS, and databases. Most data comes from the HDFS. The high throughput of HDFS can be used to read massive data. After being computed, data can be stored in HDFS.

Relationship Between MapReduce and Yarn

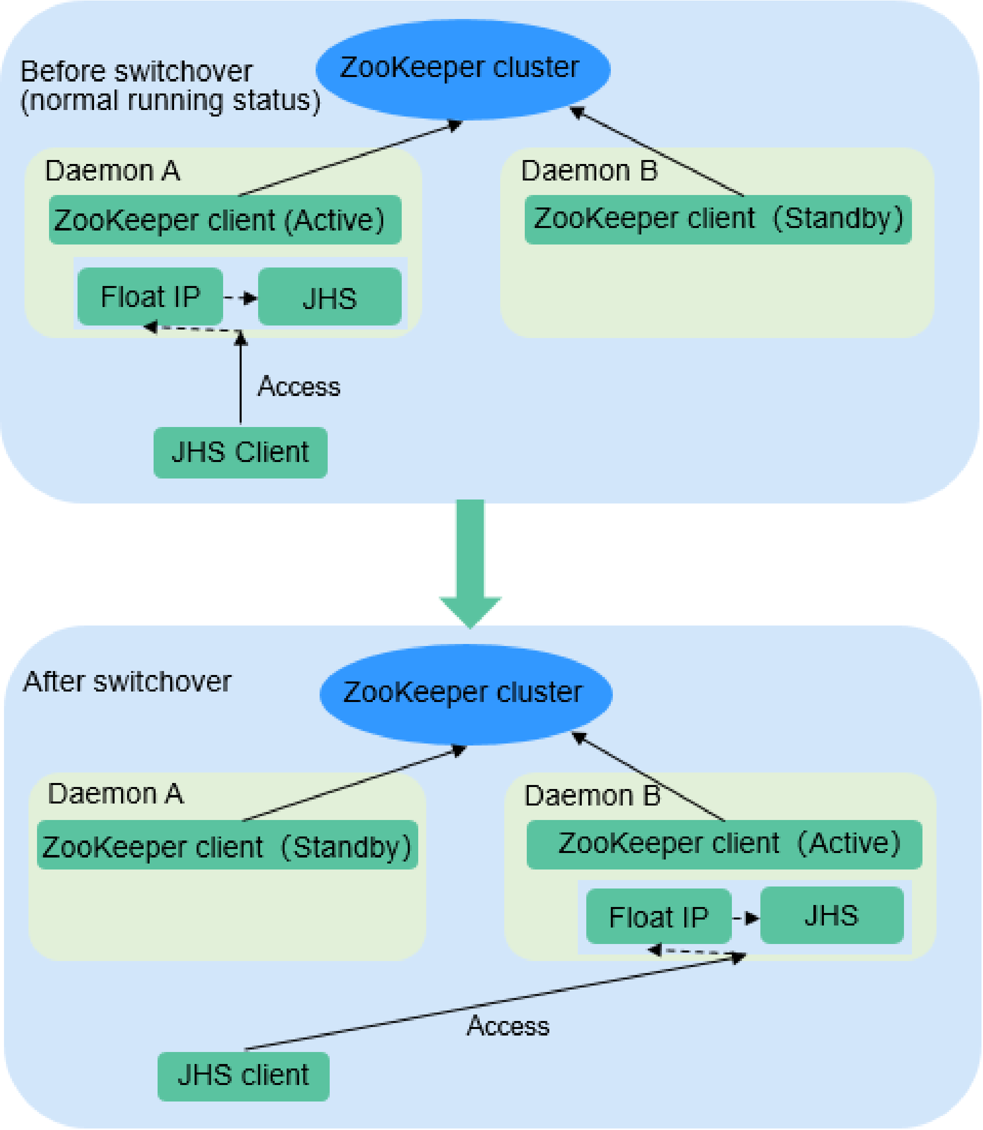
MapReduce is a computing framework running on Yarn, which is used for batch processing. MRv1 is implemented based on MapReduce in Hadoop 1.0, which is composed of programming models (new and old programming APIs), running environment (JobTracker and TaskTracker), and data processing engine (MapTask and ReduceTask). This framework is still weak in scalability, fault tolerance (JobTracker SPOF), and compatibility with multiple frameworks. (Currently, only the MapReduce computing framework is supported.) MRv2 is implemented based on MapReduce in Hadoop 2.0. The source code reuses MRv1 programming models and data processing engine implementation, and the running environment is composed of ResourceManager and ApplicationMaster. ResourceManager is a brand new resource manager system, and ApplicationMaster is responsible for cutting MapReduce job data, assigning tasks, applying for resources, scheduling tasks, and tolerating faults.

1.3.17.3 MapReduce Enhanced Open Source Features

MapReduce Enhanced Open-Source Feature: JobHistoryServer HA

JobHistoryServer (JHS) is the server used to view historical MapReduce task information. Currently, the open source JHS supports only single-instance services. JHS HA can solve the problem that an application fails to access the MapReduce API when SPOFs occur on the JHS, which causes the application fails to be executed. This greatly improves the high availability of the MapReduce service.

Figure 1-76 Status transition of the JobHistoryServer HA active/standby switchover



JobHistoryServer High Availability

- ZooKeeper is used to implement active/standby election and switchover.
- JHS uses the floating IP address to provide services externally.
- Both the JHS single-instance and HA deployment modes are supported.
- Only one node starts the JHS process at a time point to prevent multiple JHS operations from processing the same file.
- You can perform scale-out, scale-in, instance migration, upgrade, and health check.

Enhanced Open Source Feature: Improving MapReduce Performance by Optimizing the Merge/Sort Process in Specific Scenarios

The figure below shows the workflow of a MapReduce task.

Figure 1-77 MapReduce job

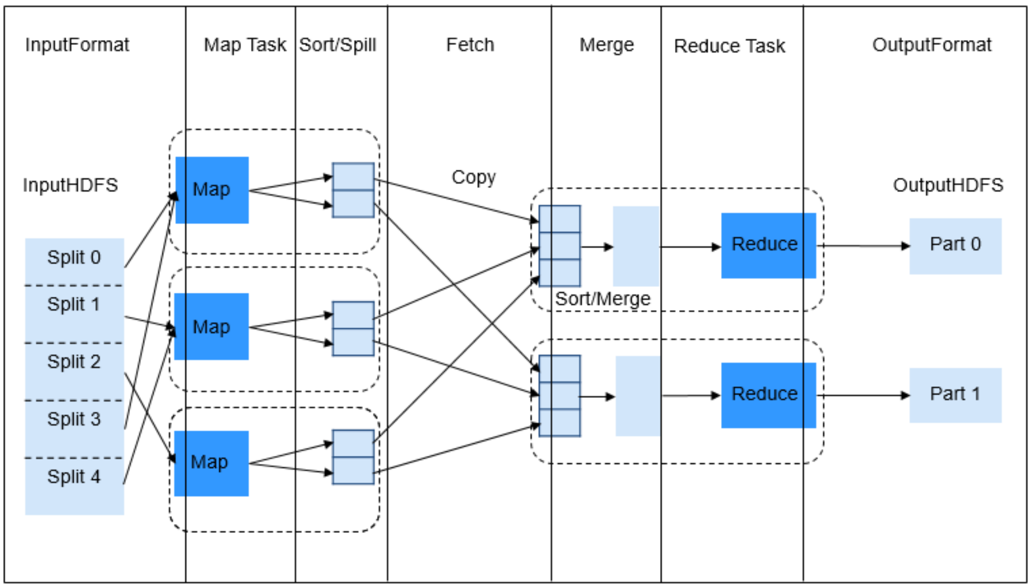
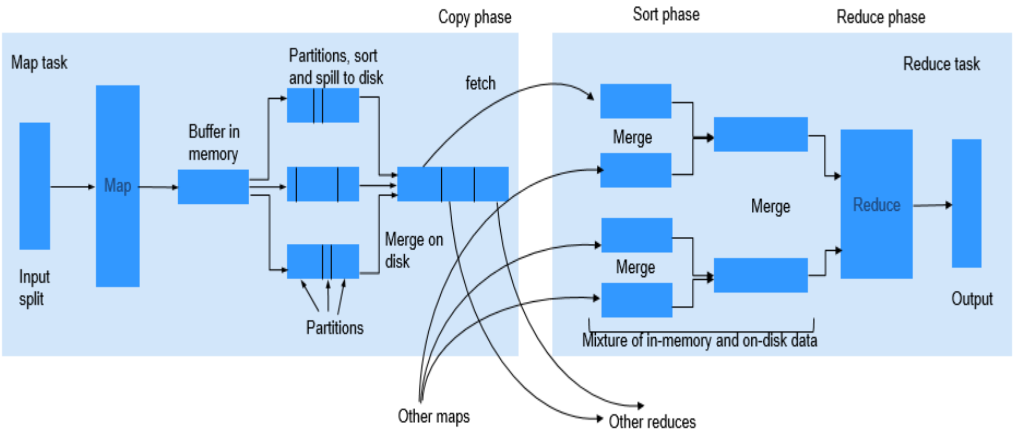


Figure 1-78 MapReduce job execution flow



The Reduce process is divided into three different steps: Copy, Sort (actually supposed to be called Merge), and Reduce. In Copy phase, Reducer tries to fetch the output of Maps from NodeManagers and store it on Reducer either in memory or on disk. Shuffle (Sort and Merge) phase then begins. All the fetched map outputs are being sorted, and segments from different map outputs are merged before being sent to Reducer. When a job has a large number of maps to be processed, the shuffle process is time-consuming. For specific tasks (for example, SQL tasks such as hash join and hash aggregation), sorting is not mandatory during the shuffle process. However, the sorting is required by default in the shuffle process.

This feature is enhanced by using the MapReduce API, which can automatically close the Sort process for such tasks. When the sorting is disabled, the API directly merges the fetched Maps output data and sends the data to Reducer. This greatly saves time, and significantly improves the efficiency of SQL tasks.

Enhanced Open Source Feature: Small Log File Problem Solved After Optimization of MR History Server

After the job running on Yarn is executed, NodeManager uses LogAggregationService to collect and send generated logs to HDFS and deletes them from the local file system. After the logs are stored to HDFS, they are managed by MR HistoryServer. LogAggregationService will merge local logs generated by containers to a log file and upload it to the HDFS, reducing the number of log files to some extent. However, in a large-scale and busy cluster, there will be excessive log files on HDFS after long-term running.

For example, if there are 20 nodes, about 18 million log files are generated within the default clean-up period (15 days), which occupy about 18 GB of the memory of a NameNode and slow down the HDFS system response.

Only the reading and deletion are required for files stored on HDFS. Therefore, Hadoop Archives can be used to periodically archive the directory of collected log files.

Archiving Logs

The AggregatedLogArchiveService module is added to MR HistoryServer to periodically check the number of files in the log directory. When the number of files reaches the threshold, AggregatedLogArchiveService starts an archiving task to archive log files. After archiving, it deletes the original log files to reduce log files on HDFS.

Cleaning Archived Logs

Hadoop Archives does not support deletion in archived files. Therefore, the entire archive log package must be deleted upon log clean-up. The latest log generation time is obtained by modifying the AggregatedLogDeletionService module. If all log files meet the clean-up requirements, the archive log package can be deleted.

Browsing Archived Logs

Hadoop Archives allows URI-based access to file content in the archive log package. Therefore, if MR History Server detects that the original log files do not exist during file browsing, it directly redirects the URI to the archive log package to access the archived log file.

NOTE

- This function invokes Hadoop Archives of HDFS for log archiving. Because the execution of an archiving task by Hadoop Archives is to run an MR application. Therefore, after an archiving task is executed, an MR execution record is added.
- This function of archiving logs is based on the log collection function. Therefore, this function is valid only when the log collection function is enabled.

1.3.18 Oozie

1.3.18.1 Oozie Basic Principles

Introduction to Oozie

Oozie is an open-source workflow engine that is used to schedule and coordinate Hadoop jobs.

Architecture

The Oozie engine is a web application integrated into Tomcat by default. Oozie uses PostgreSQL databases.

Oozie provides an Ext-based web console, through which users can view and monitor Oozie workflows. Oozie provides an external REST web service API for the Oozie client to control workflows (such as starting and stopping operations), and orchestrate and run Hadoop MapReduce tasks. For details, see **Figure 1-79**.

Figure 1-79 Oozie architecture

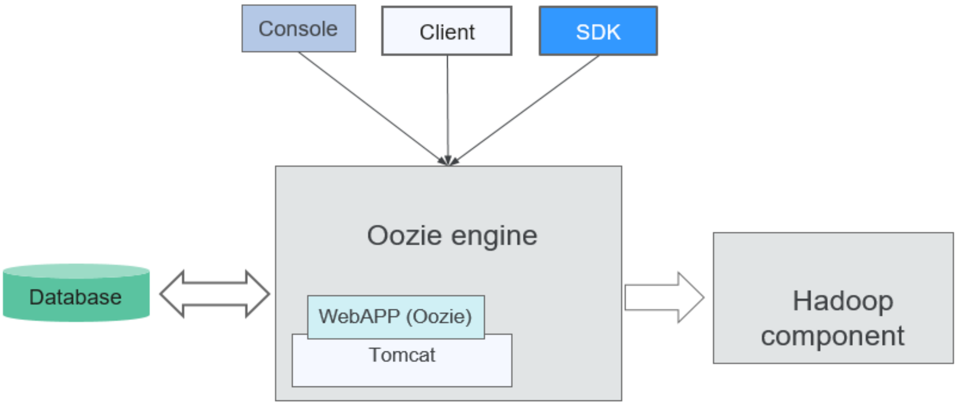


Table 1-17 describes the functions of each module shown in **Figure 1-79**.

Table 1-17 Architecture description

Connection Name	Description
Console	Allows users to view and monitor Oozie workflows.
Client	Controls workflows, including submitting, starting, running, planting, and restoring workflows, through APIs.
SDK	Is short for software development kit. An SDK is a set of development tools used by software engineers to establish applications for particular software packages, software frameworks, hardware platforms, and operating systems.
Database	PostgreSQL database

Connection Name	Description
WebApp (Oozie)	Functions as the Oozie server. It can be deployed on a built-in or an external Tomcat container. Information recorded by WebApp (Oozie) including logs is stored in the PostgreSQL database.
Tomcat	A free open-source web application server
Hadoop components	Underlying components, such as MapReduce and Hive, that execute the workflows orchestrated by Oozie.

Principle

Oozie is a workflow engine server that runs MapReduce workflows. It is also a Java web application running in a Tomcat container.

Oozie workflows are constructed using Hadoop Process Definition Language (HPDL). HPDL is an XML-defined language, similar to JBoss jBPM Process Definition Language (jPDL). An Oozie workflow consists of the Control Node and Action Node.

- Control Node controls workflow orchestration, such as **start**, **end**, **error**, **decision**, **fork**, and **join**.
- An Oozie workflow contains multiple Action Nodes, such as MapReduce and Java.

  All Action Nodes are deployed and run in Direct Acyclic Graph (DAG) mode. Therefore, Action Nodes run in direction. That is, the next Action Node can run only when the running of the previous Action Node ends. When one Action Node ends, the remote server calls back the Oozie interface. Then Oozie executes the next Action Node of workflow in the same manner until all Action Nodes are executed (execution failures are counted).

Oozie workflows provide various types of Action Nodes, such as MapReduce, Hadoop distributed file system (HDFS), Secure Shell (SSH), Java, and Oozie sub-flows, to support a wide range of business requirements.

1.3.18.2 Oozie Enhanced Open Source Features

Enhanced Open Source Feature: Improved Security

Provides roles of administrator and common users to support Oozie permission management.

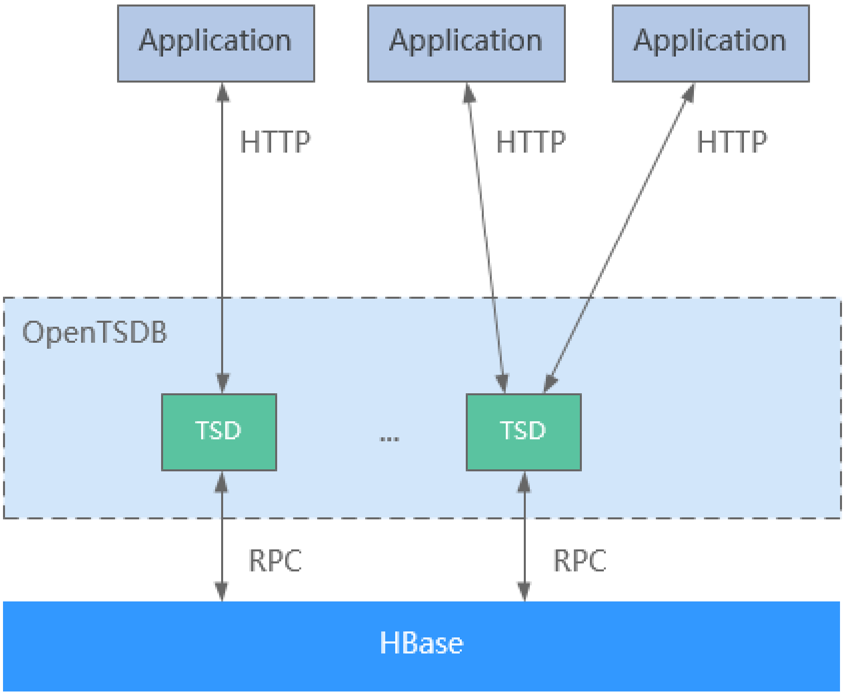
Supports single sign-on and sign-out, HTTPS access, and audit logs.

1.3.19 OpenTSDB

OpenTSDB is a distributed, scalable time series database based on HBase. OpenTSDB is designed to collect monitoring information of a large-scale cluster and implement second-level data query, eliminating the limitations of querying and storing massive amounts of monitoring data in common databases.

OpenTSDB consists of a Time Series Daemon (TSD) as well as a set of command line utilities. Interaction with OpenTSDB is primarily implemented by running one or more TSDs. Each TSD is independent. There is no master server and no shared state, so you can run as many TSDs as required to handle any load you throw at it. Each TSD uses HBase in a CloudTable cluster to store and retrieve time series data. The data schema is highly optimized for fast aggregations of similar time series to minimize storage space. TSD users never need to directly access the underlying storage. You can communicate with the TSD through an HTTP API. All communications happen on the same port (the TSD figures out the protocol of the client by looking at the first few bytes it receives).

Figure 1-80 OpenTSDB architecture



Application scenarios of OpenTSDB have the following features:

- The collected metrics have a unique value at a time point and do not have a complex structure or relationship.
- Monitoring metrics change with time.
- Like HBase, OpenTSDB features high throughput and good scalability.

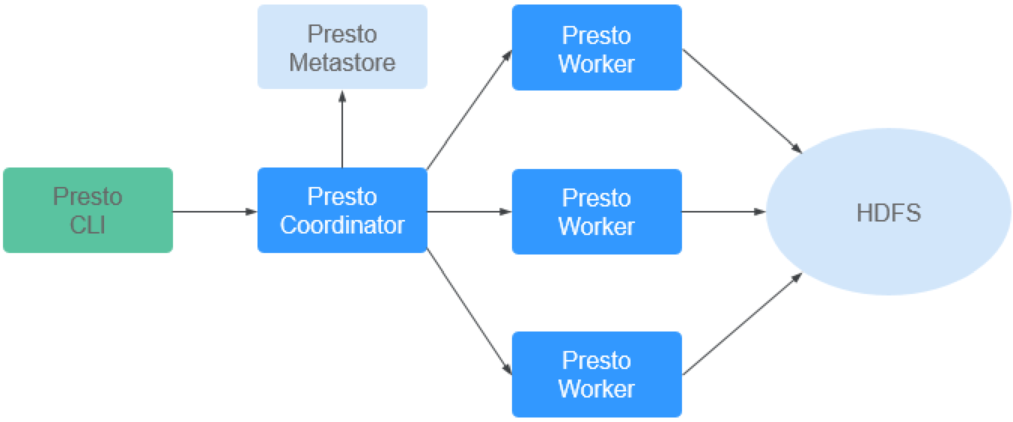
OpenTSDB provides an HTTP based application programming interface to enable integration with external systems. Almost all OpenTSDB features are accessible via the API such as querying time series data, managing metadata, and storing data points. For details, visit **https://opentsdb.net/docs/build/html/api_http/index.html**.

1.3.20 Presto

Presto is an open source SQL query engine for running interactive analytic queries against data sources of all sizes. It applies to massive structured/semi-structured data analysis, massive multi-dimensional data aggregation/report, ETL, ad-hoc queries, and more scenarios.

Presto allows querying data where it lives, including HDFS, Hive, HBase, Cassandra, relational databases or even proprietary data stores. A Presto query can combine different data sources to perform data analysis across the data sources.

Figure 1-81 Presto architecture



Presto runs in a cluster in distributed mode and contains one coordinator and multiple worker processes. Query requests are submitted from clients (for example, CLI) to the coordinator. The coordinator parses SQL statements, generates execution plans, and distributes the plans to multiple worker processes for execution.

For details about Presto, visit **https://prestodb.github.io/** or **https://prestosql.io/**.

Multiple Presto Instances

MRS supports the installation of multiple Presto instances for a large-scale cluster by default. That is, multiple Worker instances, such as Worker1, Worker2, and Worker3, are installed on a Core/Task node. Multiple Worker instances interact with the Coordinator to execute computing tasks, greatly improving node resource utilization and computing efficiency.

Presto multi-instance applies only to the Arm architecture. Currently, a single node supports a maximum of four instances.

For more Presto deployment information, see **https://prestodb.io/docs/current/installation/deployment.html** or **https://trino.io/docs/current/installation/deployment.html**.

1.3.21 Ranger

1.3.21.1 Ranger Basic Principles

Apache Ranger offers a centralized security management framework and supports unified authorization and auditing. It manages fine grained access control over Hadoop and related components, such as Storm, HDFS, Hive, HBase, and Kafka. You can use the front-end web UI console provided by Ranger to configure policies to control users' access to these components.

Figure 1-82 shows the Ranger architecture.

Figure 1-82 Ranger structure

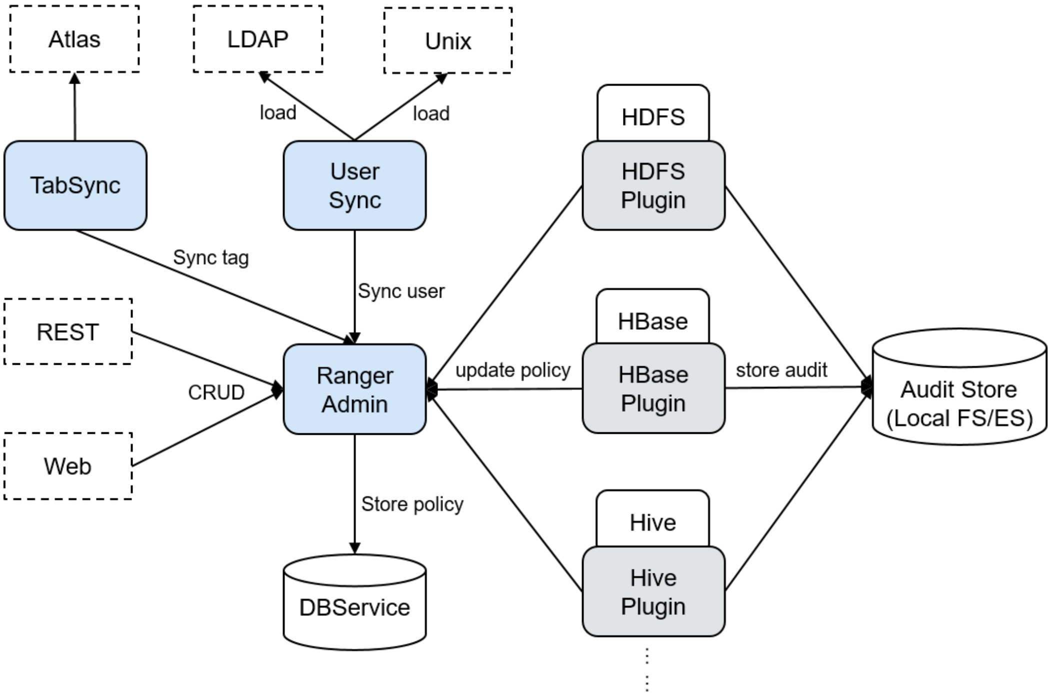


Table 1-18 Architecture description

Connection Name	Description
RangerAdmin	Provides a WebUI and RESTful API to manage policies, users, and auditing.
UserSync	Periodically synchronizes user and user group information from an external system and writes the information to RangerAdmin.
TagSync	Periodically synchronizes tag information from the external Atlas service and writes the tag information to RangerAdmin.

Ranger Principles

- Ranger Plugins

  Ranger provides policy-based access control (PBAC) plug-ins to replace the original authentication plug-ins of the components. Ranger plug-ins are developed based on the authentication interface of the components. Users set permission policies for specified services on the Ranger web UI. Ranger plug-ins periodically update policies from the RangerAdmin and caches them in the local file of the component. When a client request needs to be authenticated, the Ranger plug-in matches the user carried in the request with the policy and then returns an accept or reject message.

- UserSync User Synchronization

  UserSync periodically synchronizes data from LDAP/Unix to RangerAdmin. In security mode, data is synchronized from LDAP. In non-security mode, data is synchronized from Unix. By default, the incremental synchronization mode is used. In each synchronization period, UserSync updates only new or modified users and user groups. When a user or user group is deleted, UserSync does not synchronize the change to RangerAdmin. That is, the user or user group is not deleted from the RangerAdmin. To improve performance, UserSync does not synchronize user groups to which no user belongs to RangerAdmin.

- Unified auditing

  Ranger plug-ins can record audit logs. Currently, audit logs can be stored in local files.

- High reliability

  Ranger supports two RangerAdmins working in active/active mode. Two RangerAdmins provide services at the same time. If either RangerAdmin is faulty, Ranger continues to work.

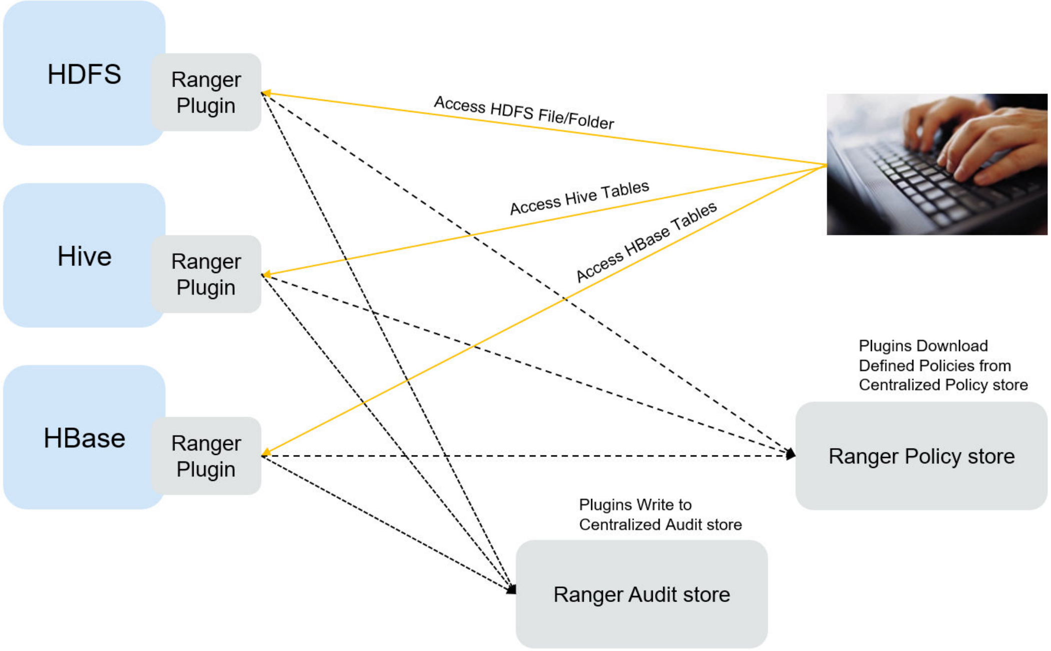
- High performance

  Ranger provides the Load-Balance capability. When a user accesses Ranger WebUI using a browser, the Load-Balance automatically selects the RangerAdmin with the lightest load to provide services.

1.3.21.2 Relationship Between Ranger and Other Components

Ranger provides PABC-based authentication plug-ins for components to run on their servers. Ranger currently supports authentication for the following components like HDFS, YARN, Hive, HBase, Kafka, Storm, and Spark2x. More components will be supported in the future.

Figure 1-83 Relationship Between Ranger and Other Components



1.3.22 Spark

1.3.22.1 Basic Principles of Spark

NOTE

The Spark component applies to versions earlier than MRS 3.x.

Description

Spark is an open source parallel data processing framework. It helps you easily develop unified big data applications and perform offline processing, stream processing, and interactive analysis on data.

Spark provides a framework featuring fast computing, write, and interactive query. Spark has obvious advantages over Hadoop in terms of performance. Spark uses the in-memory computing mode to avoid I/O bottlenecks in scenarios where multiple tasks in a MapReduce workflow process the same dataset. Spark is implemented by using Scala programming language. Scala enables distributed datasets to be processed in a method that is the same as that of processing local data. In addition to interactive data analysis, Spark supports interactive data mining. Spark adopts in-memory computing, which facilitates iterative computing. By coincidence, iterative computing of the same data is a general problem facing data mining. In addition, Spark can run in Yarn clusters where Hadoop 2.0 is installed. The reason why Spark cannot only retain various features like MapReduce fault tolerance, data localization, and scalability but also ensure high performance and avoid busy disk I/Os is that a memory abstraction structure called Resilient Distributed Dataset (RDD) is created for Spark.

Original distributed memory abstraction, for example, key-value store and databases, supports small-granularity update of variable status. This requires backup of data or log updates to ensure fault tolerance. Consequently, a large amount of I/O consumption is brought about to data-intensive workflows. For the RDD, it has only one set of restricted APIs and only supports large-granularity update, for example, map and join. In this way, Spark only needs to record the transformation operation logs generated during data establishment to ensure fault tolerance without recording a complete dataset. This data transformation link record is a source for tracing a data set. Generally, parallel applications apply the same computing process for a large dataset. Therefore, the limit to the mentioned large-granularity update is not large. As described in Spark theses, the RDD can function as multiple different computing frameworks, for example, programming models of MapReduce and Pregel. In addition, Spark allows you to explicitly make a data transformation process be persistent on hard disks. Data localization is implemented by allowing you to control data partitions based on the key value of each record. (An obvious advantage of this method is that two copies of data to be associated will be hashed in the same mode.) If memory usage exceeds the physical limit, Spark writes relatively large partitions into hard disks, thereby ensuring scalability.

Spark has the following features:

- Fast: The data processing speed of Spark is 10 to 100 times higher than that of MapReduce.
- Easy-to-use: Java, Scala, and Python can be used to simply and quickly compile parallel applications for processing massive amounts of data. Spark provides over 80 operators to help you compile parallel applications.
- Universal: Spark provides many tools, for example, **Spark SQL** and **Spark Streaming**. These tools can be combined flexibly in an application.
- Integration with Hadoop: Spark can directly run in a Hadoop cluster and read existing Hadoop data.

The Spark component of MRS has the following advantages:

- The Spark Streaming component of MRS supports real-time data processing rather than triggering as scheduled.
- The Spark component of MRS provides Structured Streaming and allows you to build streaming applications using the Dataset API. Spark supports exactly-once semantics and inner and outer joins for streams.
- The Spark component of MRS uses **pandas_udf** to replace the original user-defined functions (UDFs) in PySpark to process data, which reduces the processing duration by 60% to 90% (affected by specific operations).
- The Spark component of MRS also supports graph data processing and allows modeling using graphs during graph computing.
- Spark SQL of MRS is compatible with some Hive syntax (based on the 64 SQL statements of the Hive-Test-benchmark test set) and standard SQL syntax (based on the 99 SQL statements of the TPC-DS test set).

For details about Spark architecture and principles, visit **https://spark.apache.org/docs/3.1.1/quick-start.html**.

Architecture

Figure 1-84 describes the Spark architecture and **Table 1-19** lists the Spark modules.

Figure 1-84 Spark architecture

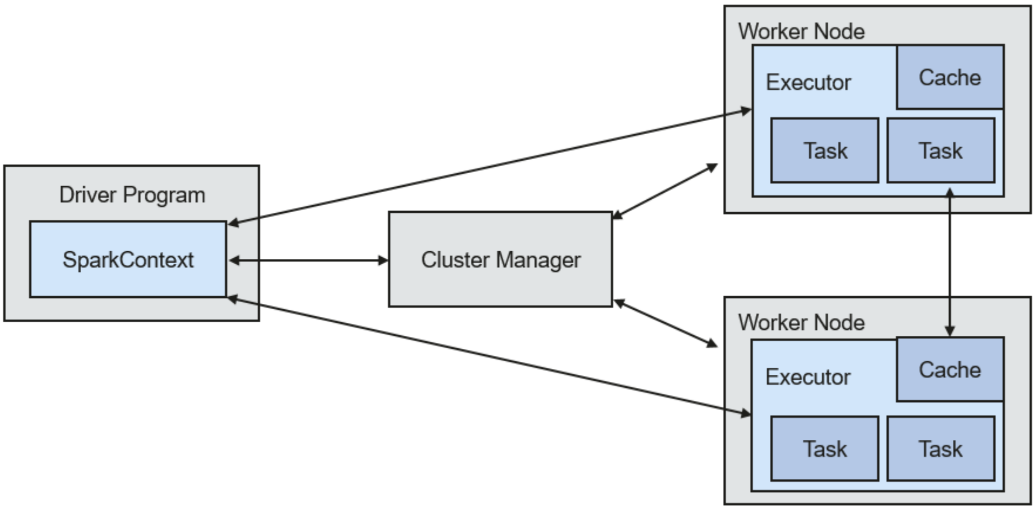


Table 1-19 Basic concepts

Module	Description
Cluster Manager	Cluster manager manages resources in the cluster. Spark supports multiple cluster managers, including Mesos, Yarn, and the Standalone cluster manager that is delivered with Spark.
Application	Spark application. It consists of one Driver Program and multiple executors.
Deploy Mode	Deployment in cluster or client mode. In cluster mode, the driver runs on a node inside the cluster. In client mode, the driver runs on the client (outside the cluster).
Driver Program	The main process of the Spark application. It runs the **main()** function of an application and creates SparkContext. It is used for parsing applications, generating stages, and scheduling tasks to executors. Usually, SparkContext represents Driver Program.
Executor	A process started on a Work Node. It is used to execute tasks, and manage and process the data used in applications. A Spark application usually contains multiple executors. Each executor receives commands from the driver and executes one or multiple tasks.
Worker Node	A node that starts and manages executors and resources in a cluster.
Job	A job consists of multiple concurrent tasks. One action operator (for example, a collect operator) maps to one job.
Stage	Each job consists of multiple stages. Each stage is a task set, which is separated by Directed Acyclic Graph (DAG).

Module	Description
Task	A task carries the computation unit of the service logics. It is the minimum working unit that can be executed on the Spark platform. An application can be divided into multiple tasks based on the execution plan and computation amount.

Spark Application Running Principle

Figure 1-85 shows the Spark application running architecture. The running process is as follows:

1. An application is running in the cluster as a collection of processes. Driver coordinates the running of the application.
2. To run an application, Driver connects to the cluster manager (such as Standalone, Mesos, and Yarn) to apply for the executor resources, and start ExecutorBackend. The cluster manager schedules resources between different applications. Driver schedules DAGs, divides stages, and generates tasks for the application at the same time.
3. Then, Spark sends the codes of the application (the codes transferred to **SparkContext**, which is defined by JAR or Python) to an executor.
4. After all tasks are finished, the running of the user application is stopped.

Figure 1-85 Spark application running architecture

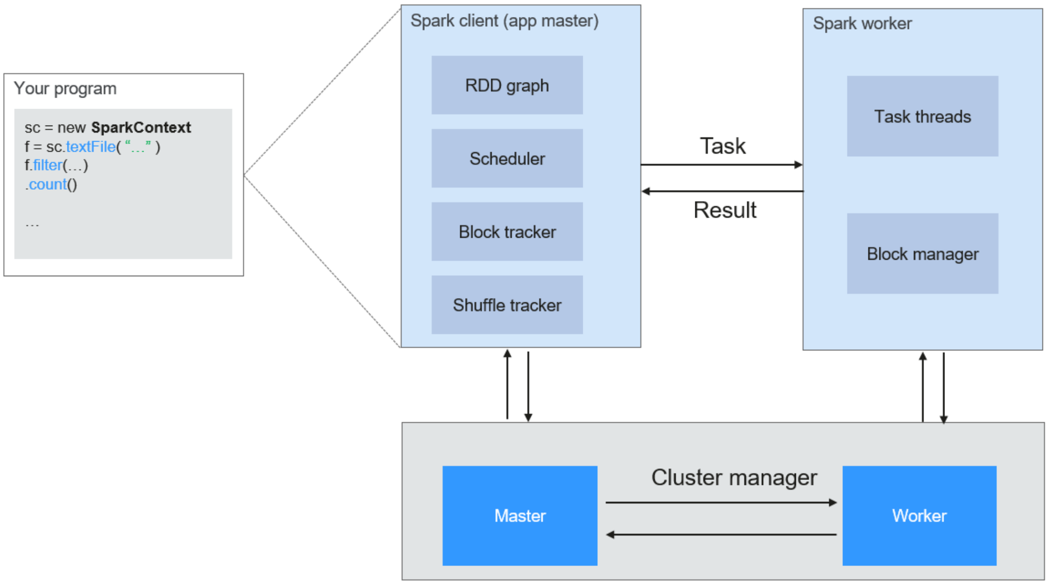
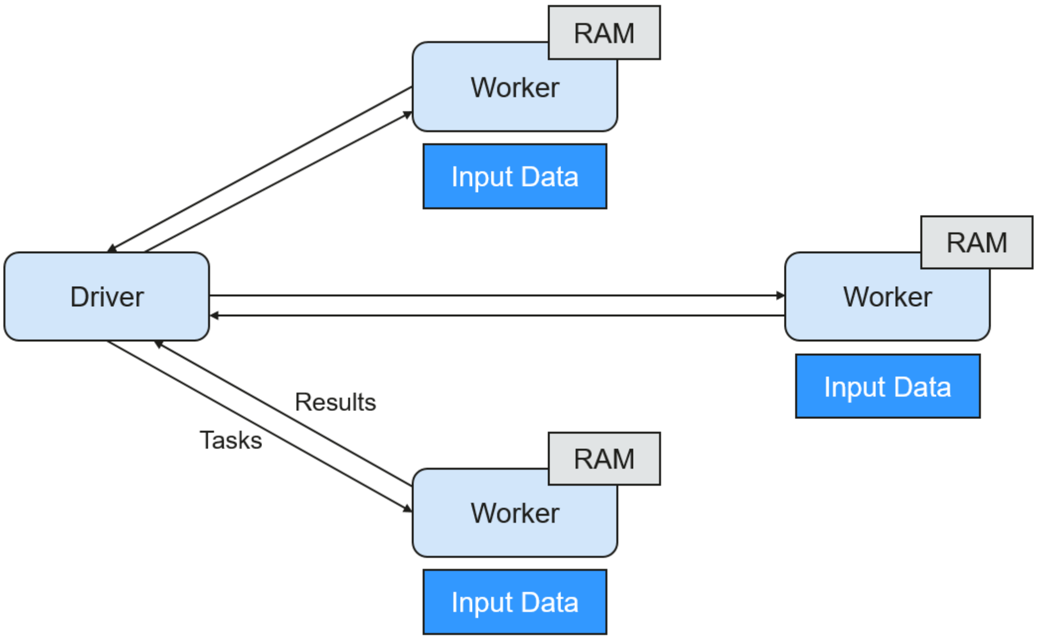


Figure 1-86 shows the Master and Worker modes adopted by Spark. A user submits an application on the Spark client, and then the scheduler divides a job into multiple tasks and sends the tasks to each Worker for execution. Each Worker reports the computation results to Driver (Master), and then the Driver aggregates and returns the results to the client.

Figure 1-86 Spark Master-Worker mode



Note the following about the architecture:

- Applications are isolated from each other.

  Each application has an independent executor process, and each executor starts multiple threads to execute tasks in parallel. Whether in terms of scheduling or task running on executors. Each driver independently schedules its own tasks. Different application tasks run on different JVMs, that is, different executors.
- Different Spark applications do not share data, unless data is stored in the external storage system such as HDFS.
- You are advised to deploy the Driver program in a location that is close to the Worker node because the Driver program schedules tasks in the cluster. For example, deploy the Driver program on the network where the Worker node is located.

Spark on YARN can be deployed in two modes:

- In Yarn-cluster mode, the Spark driver runs inside an ApplicationMaster process which is managed by Yarn in the cluster. After the ApplicationMaster is started, the client can exit without interrupting service running.
- In Yarn-client mode, the driver is started in the client process, and the ApplicationMaster process is used only to apply for resources from the Yarn cluster.

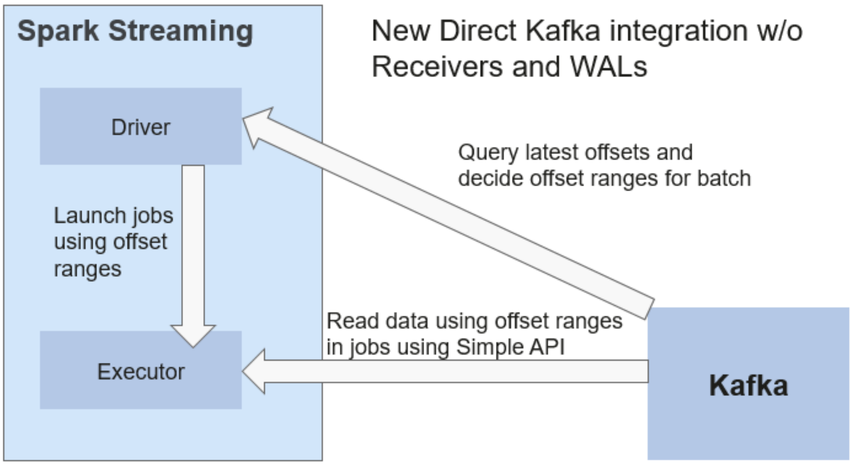
Spark Streaming Principle

Spark Streaming is a real-time computing framework built on the Spark, which expands the capability for processing massive streaming data. Currently, Spark supports the following data processing methods:

- Direct Streaming

In Direct Streaming approach, Direct API is used to process data. Take Kafka Direct API as an example. Direct API provides offset location that each batch range will read from, which is much simpler than starting a receiver to continuously receive data from Kafka and written data to write-ahead logs (WALs). Then, each batch job is running and the corresponding offset data is ready in Kafka. These offset information can be securely stored in the checkpoint file and read by applications that failed to start.

Figure 1-87 Data transmission through Direct Kafka API


Spark Streaming
New Direct Kafka integration w/o Receivers and WALs
Driver
Query latest offsets and decide offset ranges for batch
Launch jobs using offset ranges
Read data using offset ranges in jobs using Simple API
Executor
Kafka

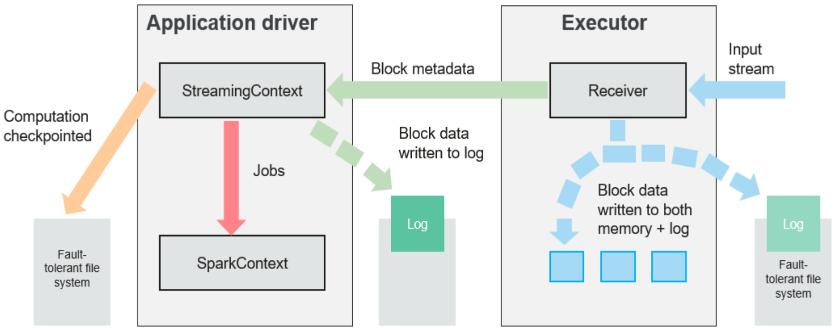
After the failure, Spark Streaming can read data from Kafka again and process the data segment. The processing result is the same no matter Spark Streaming fails or not, because the semantic is processed only once.

Direct API does not need to use the WAL and Receivers, and ensures that each Kafka record is received only once, which is more efficient. In this way, the Spark Streaming and Kafka can be well integrated, making streaming channels be featured with high fault-tolerance, high efficiency, and ease-of-use. Therefore, you are advised to use Direct Streaming to process data.

- Receiver

  When a Spark Streaming application starts (that is, when the driver starts), the related StreamingContext (the basis of all streaming functions) uses SparkContext to start the receiver to become a long-term running task. These receivers receive and save streaming data to the Spark memory for processing. Figure 1-88 shows the data transfer lifecycle.

Figure 1-88 Data transfer lifecycle



a. Receive data (blue arrow).

   Receiver divides a data stream into a series of blocks and stores them in the executor memory. In addition, after WAL is enabled, it writes data to the WAL of the fault-tolerant file system.

b. Notify the driver (green arrow).

   The metadata in the received block is sent to StreamingContext in the driver. The metadata includes:

   - Block reference ID used to locate the data position in the Executor memory.
   - Block data offset information in logs (if the WAL function is enabled).

c. Process data (red arrow).

   For each batch of data, StreamingContext uses block information to generate resilient distributed datasets (RDDs) and jobs. StreamingContext executes jobs by running tasks to process blocks in the executor memory.

d. Periodically set checkpoints (orange arrows).

   For fault tolerance, StreamingContext periodically sets checkpoints and saves them to external file systems.

Fault Tolerance

Spark and its RDD allow seamless processing of failures of any Worker node in the cluster. Spark Streaming is built on top of Spark. Therefore, the Worker node of Spark Streaming also has the same fault tolerance capability. However, Spark Streaming needs to run properly in case of long-time running. Therefore, Spark must be able to recover from faults through the driver process (main process that coordinates all Workers). This poses challenges to the Spark driver fault-tolerance because the Spark driver may be any user application implemented in any computation mode. However, Spark Streaming has internal computation architecture. That is, it periodically executes the same Spark computation in each batch data. Such architecture allows it to periodically store checkpoints to reliable storage space and recover them upon the restart of Driver.

For source data such as files, the Driver recovery mechanism can ensure zero data loss because all data is stored in a fault-tolerant file system such as HDFS. However, for other data sources such as Kafka and Flume, some received data is cached only in memory and may be lost before being processed. This is caused by the distribution operation mode of Spark applications. When the driver process fails, all executors running in the Cluster Manager, together with all data in the memory, are terminated. To avoid such data loss, the WAL function is added to Spark Streaming.

WAL is often used in databases and file systems to ensure persistence of any data operation. That is, first record an operation to a persistent log and perform this operation on data. If the operation fails, the system is recovered by reading the log and re-applying the preset operation. The following describes how to use WAL to ensure persistence of received data:

Receiver is used to receive data from data sources such as Kafka. As a long-time running task in Executor, Receiver receives data, and also confirms received data if supported by data sources. Received data is stored in the Executor memory, and Driver delivers a task to Executor for processing.

After WAL is enabled, all received data is stored to log files in the fault-tolerant file system. Therefore, the received data does not lose even if Spark Streaming fails. Besides, receiver checks correctness of received data only after the data is pre-written into logs. Data that is cached but not stored can be sent again by data sources after the driver restarts. These two mechanisms ensure zero data loss. That is, all data is recovered from logs or re-sent by data sources.

To enable the WAL function, perform the following operations:

- Set **streamingContext.checkpoint** to configure the checkpoint directory, which is an HDFS file path used to store streaming checkpoints and WALs.
- Set **spark.streaming.receiver.writeAheadLog.enable** of SparkConf to **true** (the default value is **false**).

After WAL is enabled, all receivers have the advantage of recovering from reliable received data. You are advised to disable the multi-replica mechanism because the fault-tolerant file system of WAL may also replicate the data.

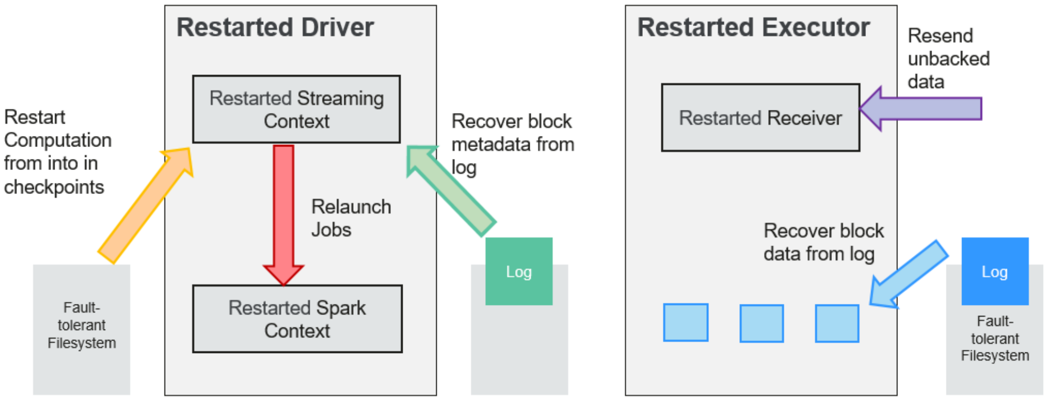
NOTE

The data receiving throughput is lowered after WAL is enabled. All data is written into the fault-tolerant file system. As a result, the write throughput of the file system and the network bandwidth for data replication may become the potential bottleneck. To solve this problem, you are advised to create more receivers to increase the degree of data receiving parallelism or use better hardware to improve the throughput of the fault-tolerant file system.

Recovery Process

When a failed driver is restarted, restart it as follows:

Figure 1-89 Computing recovery process


Restarted Driver
Restarted Streaming Context
Restarted Spark Context
Relaunch Jobs
Restart Computation from into in checkpoints
Fault-tolerant Filesystem
Recover block metadata from log
Log
Restarted Executor
Restarted Receiver
Resend unbacked data
Recover block data from log
Log
Fault-tolerant Filesystem

1. Recover computing. (Orange arrow)

   Use checkpoint information to restart Driver, reconstruct SparkContext and restart Receiver.
2. Recover metadata block. (Green arrow)

   This operation ensures that all necessary metadata blocks are recovered to continue the subsequent computing recovery.
3. Relaunch unfinished jobs. (Red arrow)

   Recovered metadata is used to generate RDDs and corresponding jobs for interrupted batch processing due to failures.

4. Read block data saved in logs. (Blue arrow)

   Block data is directly read from WALs during execution of the preceding jobs, and therefore all essential data reliably stored in logs is recovered.

5. Resend unconfirmed data. (Purple arrow)

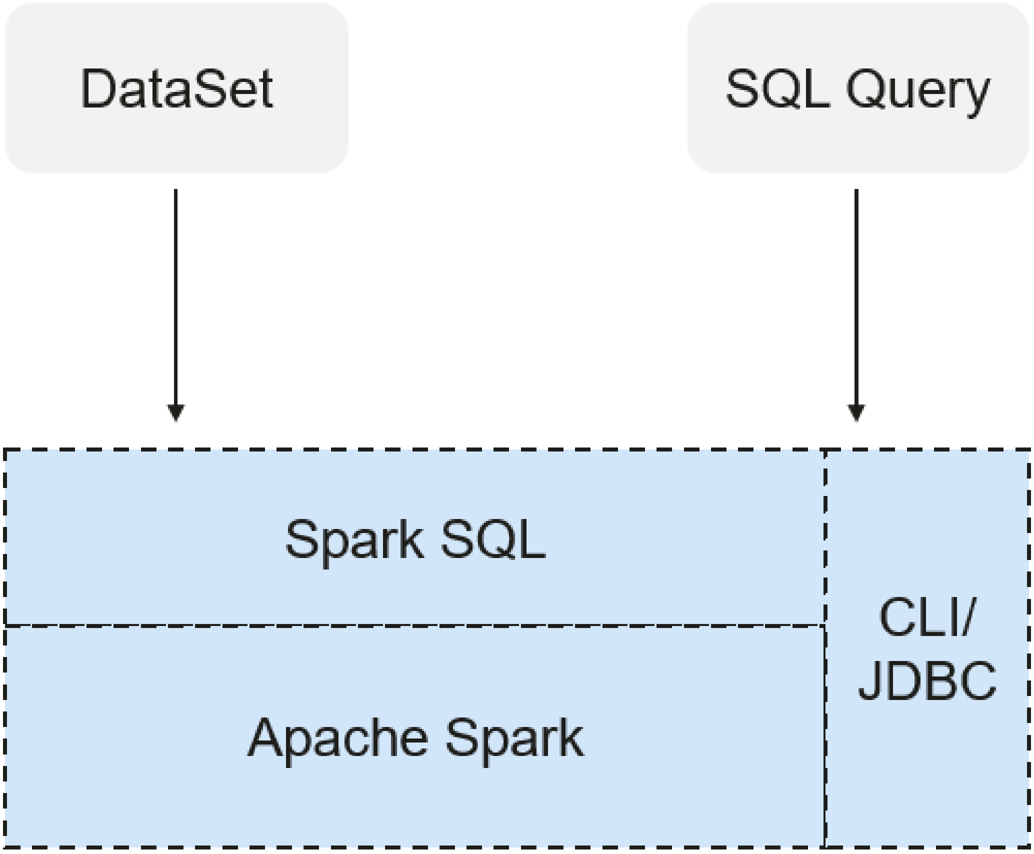
   Data that is cached but not stored to logs upon failures is re-sent by data sources, because the receiver does not confirm the data.

Therefore, by using WALs and reliable Receiver, Spark Streaming can avoid input data loss caused by Driver failures.

SparkSQL and DataSet Principle

SparkSQL

Figure 1-90 SparkSQL and DataSet



Spark SQL is a module for processing structured data. In Spark application, SQL statements or DataSet APIs can be seamlessly used for querying structured data.

Spark SQL and DataSet also provide a universal method for accessing multiple data sources such as Hive, CSV, Parquet, ORC, JSON, and JDBC. These data sources also allow data interaction. Spark SQL reuses the Hive frontend processing logic and metadata processing module. With the Spark SQL, you can directly query existing Hive data.

In addition, Spark SQL also provides API, CLI, and JDBC APIs, allowing diverse accesses to the client.

Spark SQL Native DDL/DML

In Spark 1.5, lots of Data Definition Language (DDL)/Data Manipulation Language (DML) commands are pushed down to and run on the Hive, causing coupling with the Hive and inflexibility such as unexpected error reports and results.

Spark 3.1.1 realizes command localization and replaces the Hive with Spark SQL Native DDL/DML to run DDL/DML commands. Additionally, the decoupling from the Hive is realized and commands can be customized.

DataSet

A DataSet is a strongly typed collection of domain-specific objects that can be transformed in parallel using functional or relational operations. Each Dataset also has an untyped view called a DataFrame, which is a Dataset of Row.

The DataFrame is a structured and distributed dataset consisting of multiple columns. The DataFrame is equal to a table in the relationship database or the DataFrame in the R/Python. The DataFrame is the most basic concept in the Spark SQL, which can be created by using multiple methods, such as the structured dataset, Hive table, external database or RDD.

Operations available on DataSets are divided into transformations and actions.

- A transformation operation can generate a new DataSet,

  for example, **map**, **filter**, **select**, and **aggregate (groupBy)**.
- An action operation can trigger computation and return results,

  for example, **count**, **show**, or write data to the file system.

You can use either of the following methods to create a DataSet:

- The most common way is by pointing Spark to some files on storage systems, using the **read** function available on a SparkSession.
  ```
  val people = spark.read.parquet("...").as[Person]  // Scala
  DataSet<Person> people = spark.read().parquet("...").as(Encoders.bean(Person.class));//Java
  ```
- You can also create a DataSet using the transformation operation available on an existing one.

  For example, apply the map operation on an existing DataSet to create a DataSet:
  ```
  val names = people.map(_.name) // In Scala: names is Dataset.
  Dataset<String> names = people.map((Person p) -> p.name, Encoders.STRING)); // Java
  ```

CLI and JDBCServer

In addition to programming APIs, Spark SQL also provides the CLI/JDBC APIs.

- Both **spark-shell** and **spark-sql** scripts can provide the CLI for debugging.
- JDBCServer provides JDBC APIs. External systems can directly send JDBC requests to calculate and parse structured data.

SparkSession Principle

SparkSession is a unified API for Spark programming and can be regarded as a unified entry for reading data. SparkSession provides a single entry point to perform many operations that were previously scattered across multiple classes, and also provides accessor methods to these older classes to maximize compatibility.

A SparkSession can be created using a builder pattern. The builder will automatically reuse the existing SparkSession if there is a SparkSession; or create a SparkSession if it does not exist. During I/O transactions, the configuration item settings in the builder are automatically synchronized to Spark and Hadoop.

```
import org.apache.spark.sql.SparkSession
val sparkSession = SparkSession.builder
  .master("local")
  .appName("my-spark-app")
  .config("spark.some.config.option", "config-value")
  .getOrCreate()
```

- SparkSession can be used to execute SQL queries on data and return results as DataFrame.

  ```
  sparkSession.sql("select * from person").show
  ```

- SparkSession can be used to set configuration items during running. These configuration items can be replaced with variables in SQL statements.

  ```
  sparkSession.conf.set("spark.some.config", "abcd")
  sparkSession.conf.get("spark.some.config")
  sparkSession.sql("select ${spark.some.config}")
  ```

- SparkSession also includes a "catalog" method that contains methods to work with Metastore (data catalog). After this method is used, a dataset is returned, which can be run using the same Dataset API.

  ```
  val tables = sparkSession.catalog.listTables()
  val columns = sparkSession.catalog.listColumns("myTable")
  ```

- Underlying SparkContext can be accessed by SparkContext API of SparkSession.

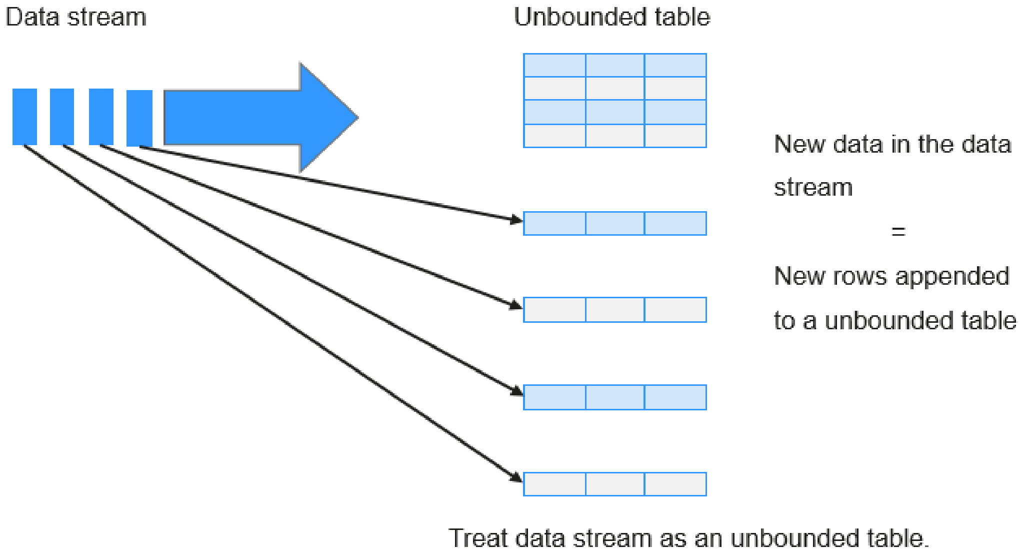
  ```
  val sparkContext = sparkSession.sparkContext
  ```

Structured Streaming Principle

Structured Streaming is a stream processing engine built on the Spark SQL engine. You can use the Dataset/DataFrame API in Scala, Java, Python, or R to express streaming aggregations, event-time windows, and stream-stream joins. If streaming data is incrementally and continuously produced, Spark SQL will continue to process the data and synchronize the result to the result set. In addition, the system ensures end-to-end exactly-once fault-tolerance guarantees through checkpoints and WALs.

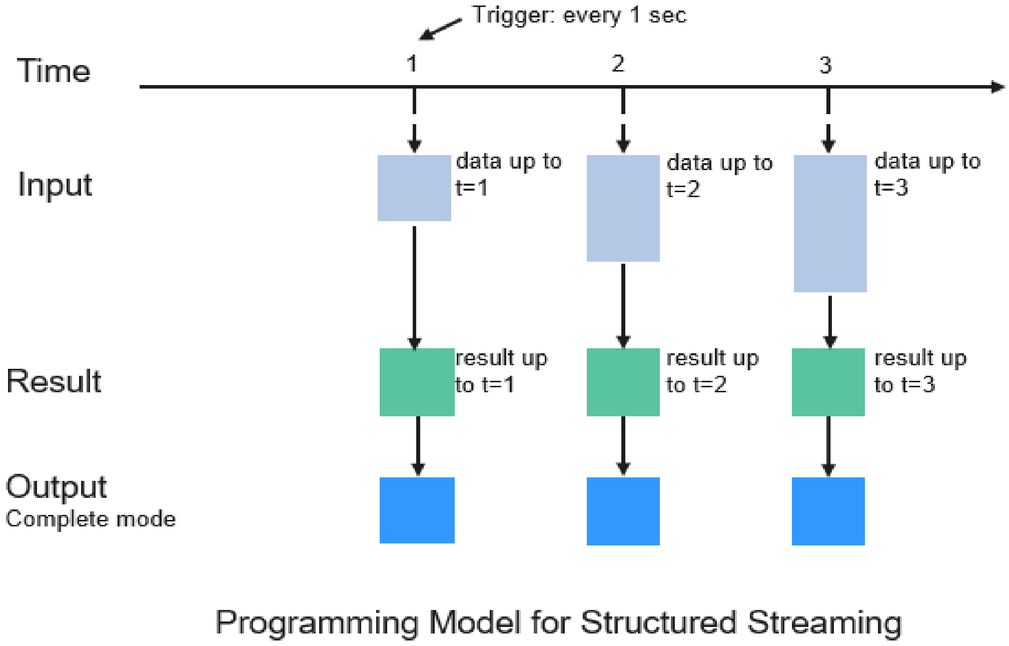
The core of Structured Streaming is to take streaming data as an incremental database table. Similar to the data block processing model, the streaming data processing model applies query operations on a static database table to streaming computing, and Spark uses standard SQL statements for query, to obtain data from the incremental and unbounded table.

Figure 1-91 Unbounded table of Structured Streaming


Data stream
Unbounded table
New data in the data stream
=
New rows appended to a unbounded table
Treat data stream as an unbounded table.

Each query operation will generate a result table. At each trigger interval, updated data will be synchronized to the result table. Whenever the result table is updated, the updated result will be written into an external storage system.

Figure 1-92 Structured Streaming data processing model



Storage modes of Structured Streaming at the output phase are as follows:

- Complete Mode: The updated result sets are written into the external storage system. The write operation is performed by a connector of the external storage system.

- Append Mode: If an interval is triggered, only added data in the result table will be written into an external system. This is applicable only on the queries where existing rows in the result table are not expected to change.
- Update Mode: If an interval is triggered, only updated data in the result table will be written into an external system, which is the difference between the Complete Mode and Update Mode.

Basic Concepts

- **RDD**

  Resilient Distributed Dataset (RDD) is a core concept of Spark. It indicates a read-only and partitioned distributed dataset. Partial or all data of this dataset can be cached in the memory and reused between computations.

  RDD Creation

  - An RDD can be created from the input of HDFS or other storage systems that are compatible with Hadoop.
  - A new RDD can be converted from a parent RDD.
  - An RDD can be converted from a collection of datasets through encoding.

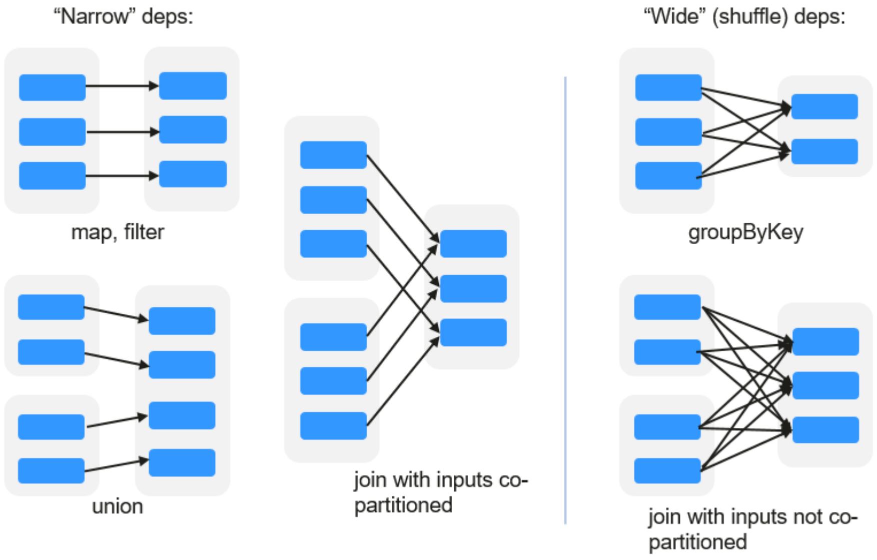
  RDD Storage

  - You can select different storage levels to store an RDD for reuse. (There are 11 storage levels to store an RDD.)
  - By default, the RDD is stored in the memory. When the memory is insufficient, the RDD overflows to the disk.

- **RDD Dependency**

  The RDD dependency includes the narrow dependency and wide dependency.

  Figure 1-93 RDD dependency


"Narrow" deps:
"Wide" (shuffle) deps:
map, filter
groupByKey
union
join with inputs co-partitioned
join with inputs not co-partitioned

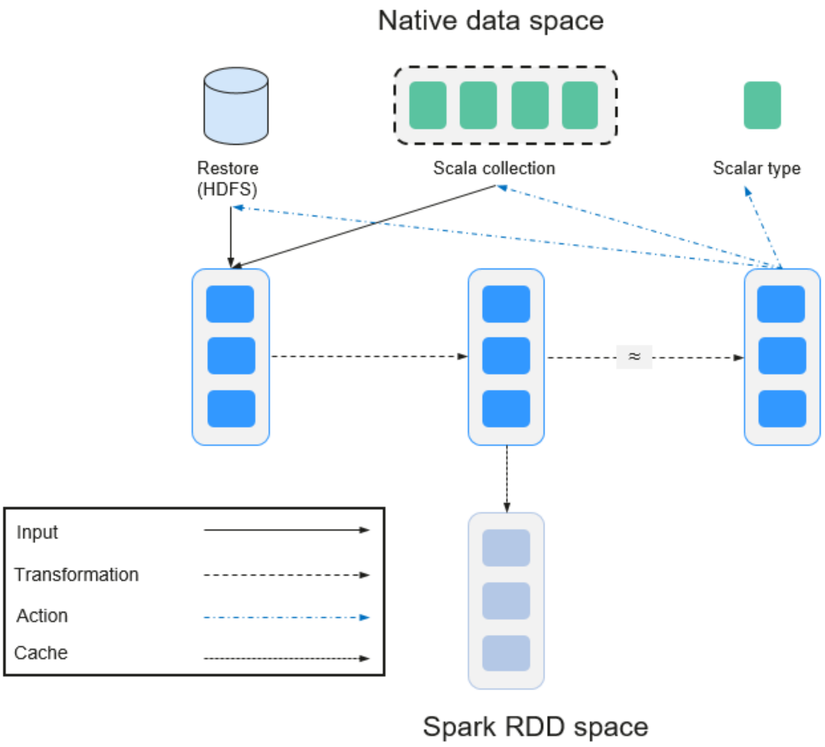
  - **Narrow dependency**: Each partition of the parent RDD is used by at most one partition of the child RDD.
  - **Wide dependency**: Partitions of the child RDD depend on all partitions of the parent RDD.

The narrow dependency facilitates the optimization. Logically, each RDD operator is a fork/join (the join is not the join operator mentioned above but the barrier used to synchronize multiple concurrent tasks); fork the RDD to each partition, and then perform the computation. After the computation, join the results, and then perform the fork/join operation on the next RDD operator. It is uneconomical to directly translate the RDD into physical implementation. The first is that every RDD (even intermediate result) needs to be physicalized into memory or storage, which is time-consuming and occupies much space. The second is that as a global barrier, the join operation is very expensive and the entire join process will be slowed down by the slowest node. If the partitions of the child RDD narrowly depend on that of the parent RDD, the two fork/join processes can be combined to implement classic fusion optimization. If the relationship in the continuous operator sequence is narrow dependency, multiple fork/join processes can be combined to reduce a large number of global barriers and eliminate the physicalization of many RDD intermediate results, which greatly improves the performance. This is called pipeline optimization in Spark.

- **Transformation and Action (RDD Operations)**

  Operations on RDD include transformation (the return value is an RDD) and action (the return value is not an RDD). **Figure 1-94** shows the RDD operation process. The transformation is lazy, which indicates that the transformation from one RDD to another RDD is not immediately executed. Spark only records the transformation but does not execute it immediately. The real computation is started only when the action is started. The action returns results or writes the RDD data into the storage system. The action is the driving force for Spark to start the computation.

Figure 1-94 RDD operation



The data and operation model of RDD are quite different from those of Scala.

```
val file = sc.textFile("hdfs://...")
val errors = file.filter(_.contains("ERROR"))
errors.cache()
errors.count()
```

a. The textFile operator reads log files from the HDFS and returns files (as an RDD).

b. The filter operator filters rows with **ERROR** and assigns them to errors (a new RDD). The filter operator is a transformation.

c. The cache operator caches errors for future use.

d. The count operator returns the number of rows of errors. The count operator is an action.

Transformation includes the following types:

- The RDD elements are regarded as simple elements.

  The input and output has the one-to-one relationship, and the partition structure of the result RDD remains unchanged, for example, map.

  The input and output has the one-to-many relationship, and the partition structure of the result RDD remains unchanged, for example, flatMap (one element becomes a sequence containing multiple elements after map and then flattens to multiple elements).

  The input and output has the one-to-one relationship, but the partition structure of the result RDD changes, for example, union (two RDDs integrates to one RDD, and the number of partitions becomes the sum of the number of partitions of two RDDs) and coalesce (partitions are reduced).

  Operators of some elements are selected from the input, such as filter, distinct (duplicate elements are deleted), subtract (elements only exist in this RDD are retained), and sample (samples are taken).

- The RDD elements are regarded as key-value pairs.

  Perform the one-to-one calculation on the single RDD, such as mapValues (the partition mode of the source RDD is retained, which is different from map).

  Sort the single RDD, such as sort and partitionBy (partitioning with consistency, which is important to the local optimization).

  Restructure and reduce the single RDD based on key, such as groupByKey and reduceByKey.

  Join and restructure two RDDs based on the key, such as join and cogroup.

  NOTE

  The later three operations involving sorting are called shuffle operations.

Action includes the following types:

- Generate scalar configuration items, such as **count** (the number of elements in the returned RDD), **reduce**, **fold/aggregate** (the number of scalar configuration items that are returned), and **take** (the number of elements before the return).

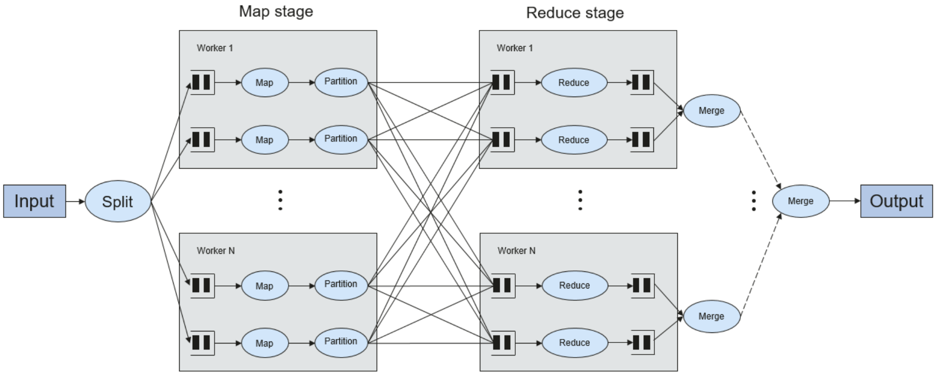
- Generate the Scala collection, such as **collect** (import all elements in the RDD to the Scala collection) and **lookup** (look up all values corresponds to the key).
- Write data to the storage, such as **saveAsTextFile** (which corresponds to the preceding **textFile**).
- Check points, such as the **checkpoint** operator. When Lineage is quite long (which occurs frequently in graphics computation), it takes a long period of time to execute the whole sequence again when a fault occurs. In this case, checkpoint is used as the check point to write the current data to stable storage.

- **Shuffle**

  Shuffle is a specific phase in the MapReduce framework, which is located between the Map phase and the Reduce phase. If the output results of Map are to be used by Reduce, the output results must be hashed based on a key and distributed to each Reducer. This process is called Shuffle. Shuffle involves the read and write of the disk and the transmission of the network, so that the performance of Shuffle directly affects the operation efficiency of the entire program.

  The figure below shows the entire process of the MapReduce algorithm.

  Figure 1-95 Algorithm process



  Shuffle is a bridge to connect data. The following describes the implementation of shuffle in Spark.

  Shuffle divides a job of Spark into multiple stages. The former stages contain one or more ShuffleMapTasks, and the last stage contains one or more ResultTasks.

- **Spark Application Structure**

  The Spark application structure includes the initialized SparkContext and the main program.

  - Initialized SparkContext: constructs the operating environment of the Spark Application.

    Constructs the SparkContext object. The following is an example:

    ```
    new SparkContext(master, appName, [SparkHome], [jars])
    ```

    Parameter description:

    master: indicates the link string. The link modes include local, Yarn-cluster, and Yarn-client.

appName: indicates the application name.

SparkHome: indicates the directory where Spark is installed in the cluster.

jars: indicates the code and dependency package of an application.

  – Main program: processes data.

  For details about how to submit an application, visit **https://spark.apache.org/docs/3.1.1/submitting-applications.html**.

- **Spark Shell Commands**

  The basic Spark shell commands support the submission of Spark applications. The Spark shell commands are as follows:

  ```
  ./bin/spark-submit \
    --class <main-class> \
    --master <master-url> \
    ... # other options
    <application-jar> \
    [application-arguments]
  ```

  Parameter description:

  --class: indicates the name of the class of a Spark application.

  --master: indicates the master to which the Spark application links, such as Yarn-client and Yarn-cluster.

  application-jar: indicates the path of the JAR file of the Spark application.

  application-arguments: indicates the parameter required to submit the Spark application. This parameter can be left blank.

- **Spark JobHistory Server**

  The Spark web UI is used to monitor the details in each phase of the Spark framework of a running or historical Spark job and provide the log display, which helps users to develop, configure, and optimize the job in more fine-grained units.

1.3.22.2 Spark HA Solution

Spark Multi-Active Instance HA Principles and Implementation Solution

Based on existing JDBCServer in the community, multi-active-instance mode is used to achieve HA. In this mode, multiple JDBCServers coexist in the cluster and the client can randomly connect any JDBCServer to perform service operations. When one or multiple JDBCServers stop working, a client can connect to another normal JDBCServer.

Compared with active/standby HA mode, multi-active instance mode has following advantages:

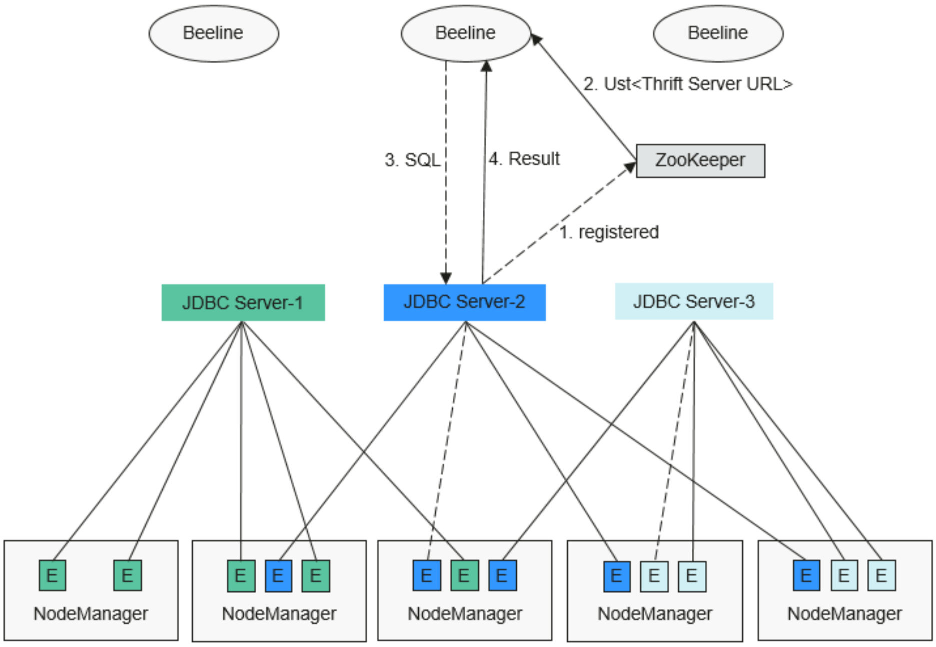
- In active/standby HA, when the active/standby switchover occurs, the unavailable period cannot be controlled by JDBCServer, but it depends on Yarn service resources.
- In Spark, the Thrift JDBC similar to HiveServer2 provides services and users access services through Beeline and JDBC API. Therefore, the processing capability of the JDBCServer cluster depends on the single-point capability of the primary server, and the scalability is insufficient.

The multi-active instance HA mode not only can prevent service interruption caused by switchover, but also enables cluster scale-out to improve high concurrency.

- **Implementation**

  The following figure shows the basic principle of multi-active instance HA of Spark JDBCServer.

Figure 1-96 Spark JDBCServer HA



1. When a JDBCServer is started, it registers with ZooKeeper by writing node information in a specified directory. Node information includes the instance IP address, port number, version, and serial number.
2. To connect to JDBCServer, the client must specify the namespace, which is the directory of JDBCServer instances in ZooKeeper. During the connection, a JDBCServer instance is randomly selected from the specified namespace.
3. After the connection succeeds, the client sends SQL statements to JDBCServer.
4. JDBCServer executes received SQL statements and returns results to the client.

If multi-active instance HA of Spark JDBCServer is enabled, all JDBCServer instances are independent and equivalent. When one JDBCServer instance is interrupted during upgrade, other JDBCServer instances can accept the connection request from the client.

The rules below must be followed in the multi-active instance HA of Spark JDBCServer.

- If a JDBCServer instance exits abnormally, no other instance will take over the sessions and services running on the abnormal instance.
- When the JDBCServer process is stopped, corresponding nodes are deleted from ZooKeeper.
- The client randomly selects the server, which may result in uneven session allocation caused by random distribution of policy results, and finally result in load imbalance of instances.

- After the instance enters the maintenance mode (in which no new connection requests from clients are accepted), services running on the instance may fail when the decommissioning times out.
- **URL Connection**
  - Multi-active instance mode

    In multi-active instance mode, the client reads content from the ZooKeeper node and connects to JDBCServer. The connection strings are list below.

    - Security mode:

      If Kinit authentication is enabled, the JDBCURL is as follows:

      ```
      jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark/hadoop.<System domain name>@<System domain name>;
      ```

      NOTE

      - In the above JDBCURL, **<zkNode_IP>:<zkNode_Port>** indicates the ZooKeeper URL. Use commas (,) to separate multiple URLs,

        Example: 192.168.81.37:2181,192.168.195.232:2181,192.168.169.84:2181.
      - **sparkthriftserver2x** indicates the ZooKeeper directory, where a random JDBCServer instance is connected to the client.

      For example, when you use Beeline client to connect JDBCServer, run the following command:

      ```
      sh CLIENT_HOME/spark/bin/beeline -u "jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark/hadoop.<System domain name>@<System domain name>;"
      ```

      If Keytab authentication is enabled, the JDBCURL is as follows:

      ```
      jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark/hadoop.<System domain name>@<System domain name>;user.principal=<principal_name>;user.keytab=<path_to_keytab>
      ```

      In the above URL, *<principal_name>* indicates the principal of the Kerberos user, for example, **test@***<System domain name>*; *<path_to_keytab>* indicates the Keytab file path corresponding to *<principal_name>*, for example, **/opt/auth/test/user.keytab**.

    - Common mode:

      ```
      jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;
      ```

      For example, when you use Beeline client, in normal mode, for connection, run the following command:

      ```
      sh CLIENT_HOME/spark/bin/beeline -u "jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;"
      ```

– Non-multi-active instance mode

In this mode, a client connects to a specified JDBCServer node. Compared with multi-active instance mode, the connection string in this mode does not contain **serviceDiscoveryMode** and **zooKeeperNamespace** parameters about ZooKeeper.

For example, when you use Beeline client, in security mode, to connect JDBCServer in non-multi-active instance mode, run the following command:

sh *CLIENT_HOME***/spark/bin/beeline -u "jdbc:hive2://** *<server_IP>:<server_Port>***/;user.principal=spark/hadoop.***<System domain name>@<System domain name>***;saslQop=auth-conf;auth=KERBEROS;principal=spark/hadoop.***<System domain name>@<System domain name>***;"**

NOTE

- In the above command, **<server_IP>:<server_Port>** indicates the URL of the specified JDBCServer node.
- **CLIENT_HOME** indicates the client path.

Except the connection method, other operations of JDBCServer API in the two modes are the same. Spark JDBCServer is another implementation of HiveServer2 in Hive. For details about how to use Spark JDBCServer, see **https://cwiki.apache.org/confluence/display/Hive/HiveServer2+Clients**.

Spark Multi-Tenant HA

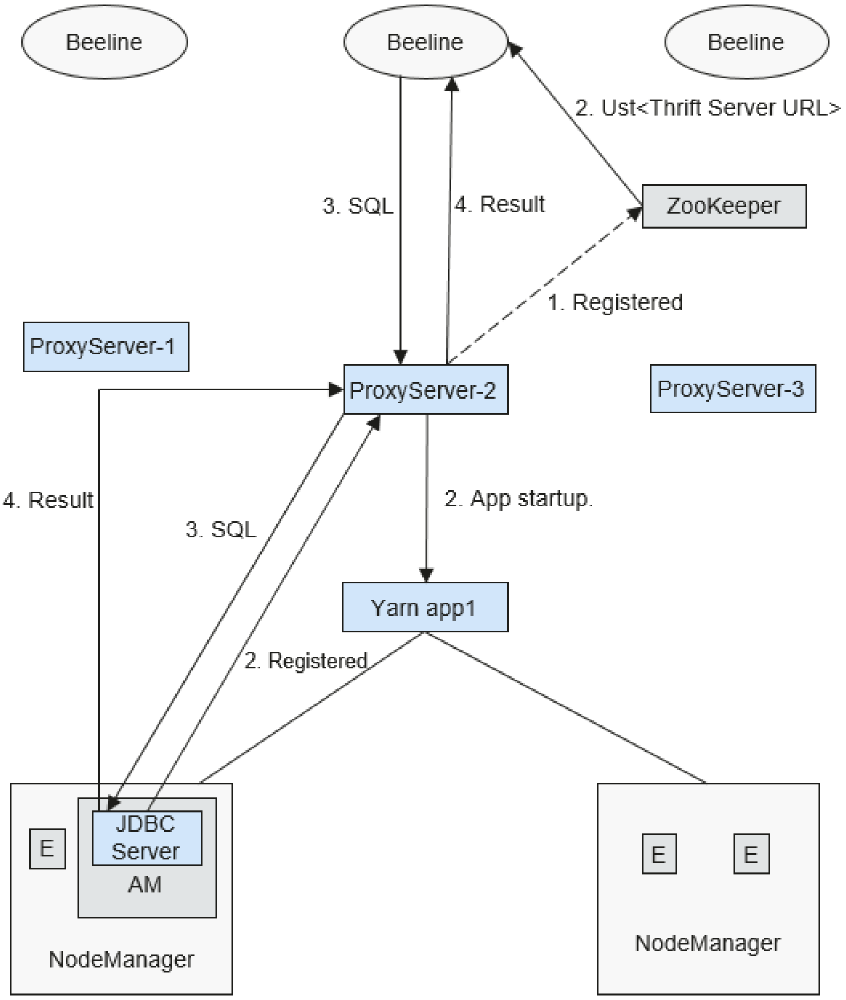
In the JDBCServer multi-active instance solution, JDBCServer uses the Yarn-client mode, but there is only one Yarn resource queue available. To solve this resource limitation problem, the multi-tenant mode is introduced.

In multi-tenant mode, JDBCServers are bound with tenants. Each tenant corresponds to one or more JDBCServers, and a JDBCServer provides services for only one tenant. Different tenants can be configured with different Yarn queues to implement resource isolation. In addition, JDBCServer can be dynamically started as required to avoid resource waste.

- **Implementation**

  Figure 1-97 shows the HA solution of the multi-tenant mode.

Figure 1-97 Multi-tenant mode of Spark JDBCServer



a. When ProxyServer is started, it registers with ZooKeeper by writing node information in a specified directory. Node information includes the instance IP address, port number, version, and serial number.

NOTE

In multi-tenant mode, the JDBCServer instance refers to the ProxyServer (JDBCServer proxy).

b. To connect to ProxyServer, the client must specify a namespace, which is the directory of the ProxyServer instance where you want to access ZooKeeper. When the client connects to the ProxyServer, a random instance under the namespace is selected for connection. For details about the URL, see **URL Connection Overview**.

c. After the client successfully connects to the ProxyServer, which first checks whether the JDBCServer of a tenant exists. If yes, Beeline connects the JDBCServer. If no, a new JDBCServer is started in Yarn-cluster mode. After the startup of JDBCServer, ProxyServer obtains the IP address of the JDBCServer and establishes the connection between Beeline and JDBCServer.

d. The client sends SQL statements to ProxyServer, which forwards statements to the connected JDBCServer. JDBCServer returns the results to ProxyServer, which then returns the results to the client.

In the multi-active instance HA mode, all instances are independent and equivalent. If one instance is interrupted during upgrade, other instances can accept the connection request from the client.

- **URL Connection Overview**
  - Multi-tenant mode

    In multi-tenant mode, the client reads content from the ZooKeeper node and connects to ProxyServer. The connection strings are list below.

    - Security mode:

      If Kinit authentication is enabled, the client URL is as follows:

      ```
      jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark/hadoop.<System domain name>@<System domain name>;
      ```

      NOTE

      - In the above URL, **<zkNode_IP>:<zkNode_Port>** indicates the ZooKeeper URL. Use commas (,) to separate multiple URLs,

        Example: **192.168.81.37:2181,192.168.195.232:2181,192.168.169.84:2181**.
      - **sparkthriftserver2x** indicates the ZooKeeper directory, where a random JDBCServer instance is connected to the client.

      For example, when you use Beeline client for connection, run the following command:

      ```
      sh CLIENT_HOME/spark/bin/beeline -u "jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark/hadoop.<System domain name>@<System domain name>;"
      ```

      If Keytab authentication is enabled, the URL is as follows:

      ```
      jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark/hadoop.<System domain name>@<System domain name>;user.principal=<principal_name>;user.keytab=<path_to_keytab>
      ```

      In the above URL, *<principal_name>* indicates the principal of the Kerberos user, for example, **test@***<System domain name>*; *<path_to_keytab>* indicates the Keytab file path corresponding to *<principal_name>*, for example, **/opt/auth/test/user.keytab**.

    - Common mode:

      ```
      jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;
      ```

      For example, run the following command when you use Beeline client for connection in normal mode:

      ```
      sh CLIENT_HOME/spark/bin/beeline -u "jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkN
      ```

*ode3_IP>:<zkNode3_Port>***/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;"**

- Non-multi-tenant mode

  In non-multi-tenant mode, a client connects to a specified JDBCServer node. Compared with multi-tenant instance mode, the connection string in this mode does not contain **serviceDiscoveryMode** and **zooKeeperNamespace** parameters about ZooKeeper.

  For example, when you use Beeline client to connect JDBCServer in non-multi-tenant instance mode, run the following command:

  sh *CLIENT_HOME***/spark/bin/beeline -u "jdbc:hive2://** *<server_IP>:<server_Port>***/;user.principal=spark/hadoop.***<System domain name>@<System domain name>***;saslQop=auth-conf;auth=KERBEROS;principal=spark/hadoop.***<System domain name>@<System domain name>***;"**

  NOTE

  - In the above command, **<server_IP>:<server_Port>** indicates the URL of the specified JDBCServer node.
  - **CLIENT_HOME** indicates the client path.

  Except the connection method, other operations of JDBCServer API in multi-tenant mode and non-multi-tenant mode are the same. Spark JDBCServer is another implementation of HiveServer2 in Hive. For details about how to use Spark JDBCServer, go to the official Hive website at **https://cwiki.apache.org/confluence/display/Hive/HiveServer2+Clients**.

  Specifying a Tenant

  Generally, the client submitted by a user connects to the default JDBCServer of the tenant to which the user belongs. If you want to connect the client to the JDBCServer of a specified tenant, add the **--hiveconf mapreduce.job.queuename** parameter.

  If you use Beeline client for connection, run the following command (**aaa** is the tenant name):

  beeline --hiveconf mapreduce.job.queuename=aaa -u 'jdbc:hive2://192.168.39.30:2181,192.168.40.210:2181,192.168.215.97:2181;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark/hadoop.*<System domain name>*@*<System domain name>*;'

1.3.22.3 Relationship Among Spark, HDFS, and Yarn

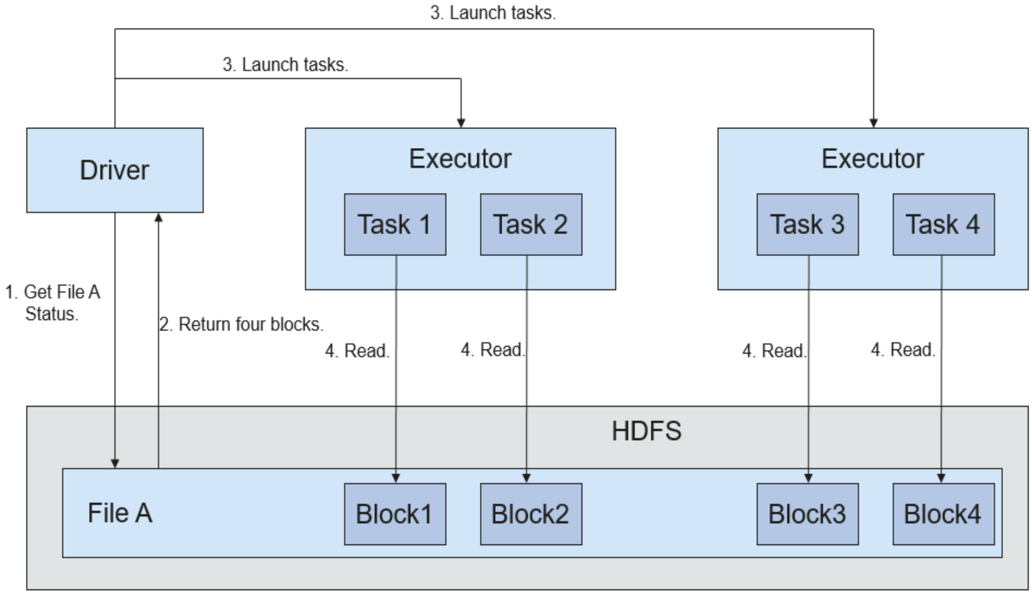
Relationship Between Spark and HDFS

Data computed by Spark comes from multiple data sources, such as local files and HDFS. Most data computed by Spark comes from the HDFS. The HDFS can read data in large scale for parallel computing. After being computed, data can be stored in the HDFS.

Spark involves Driver and Executor. Driver schedules tasks and Executor runs tasks.

Figure 1-98 shows the process of reading a file.

Figure 1-98 File reading process

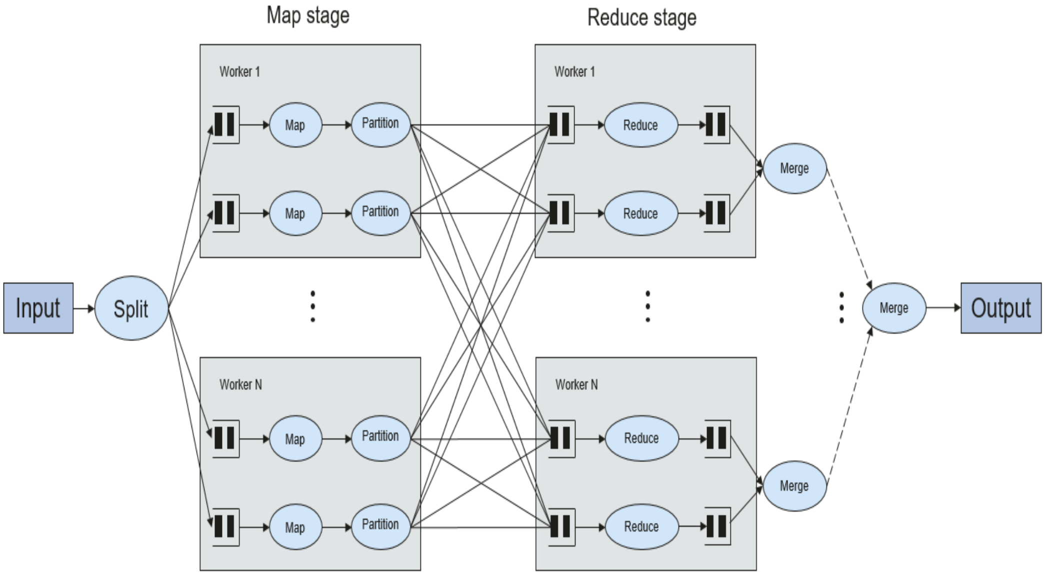


The file reading process is as follows:

1. Driver interconnects with the HDFS to obtain the information of File A.
2. The HDFS returns the detailed block information about this file.
3. Driver sets a parallel degree based on the block data amount, and creates multiple tasks to read the blocks of this file.
4. Executor runs the tasks and reads the detailed blocks as part of the Resilient Distributed Dataset (RDD).

Figure 1-99 shows the process of writing data to a file.

Figure 1-99 File writing process



The file writing process is as follows:

1. Driver creates a directory where the file is to be written.
2. Based on the RDD distribution status, the number of tasks related to data writing is computed, and these tasks are sent to Executor.
3. Executor runs these tasks, and writes the RDD data to the directory created in 1.

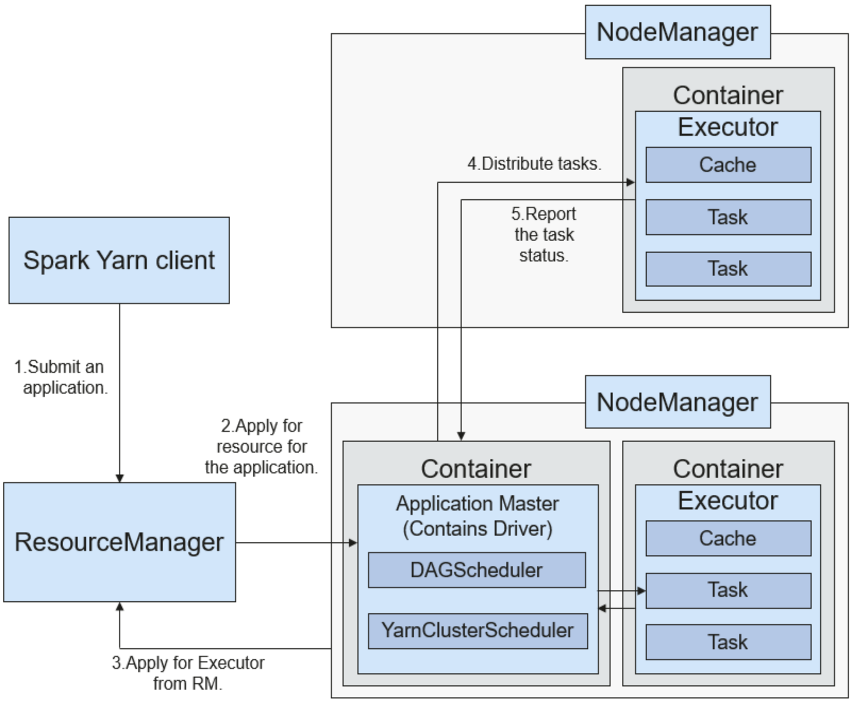
Relationship Between Spark and Yarn

The Spark computing and scheduling can be implemented using Yarn mode. Spark enjoys the computing resources provided by Yarn clusters and runs tasks in a distributed way. Spark on Yarn has two modes: Yarn-cluster and Yarn-client.

- Yarn-cluster mode

  Figure 1-100 shows the running framework of Spark on Yarn-cluster.

  Figure 1-100 Spark on Yarn-cluster operation framework



  Spark on Yarn-cluster implementation process:

  a. The client generates the application information, and then sends the information to ResourceManager.
  b. ResourceManager allocates the first container (ApplicationMaster) to SparkApplication and starts driver on the container.
  c. ApplicationMaster applies for resources from ResourceManager to run the container.

     ResourceManager allocates the container to ApplicationMaster, which communicates with NodeManager, and starts the executor in the obtained container. After the executor is started, it registers with the driver and applies for tasks.

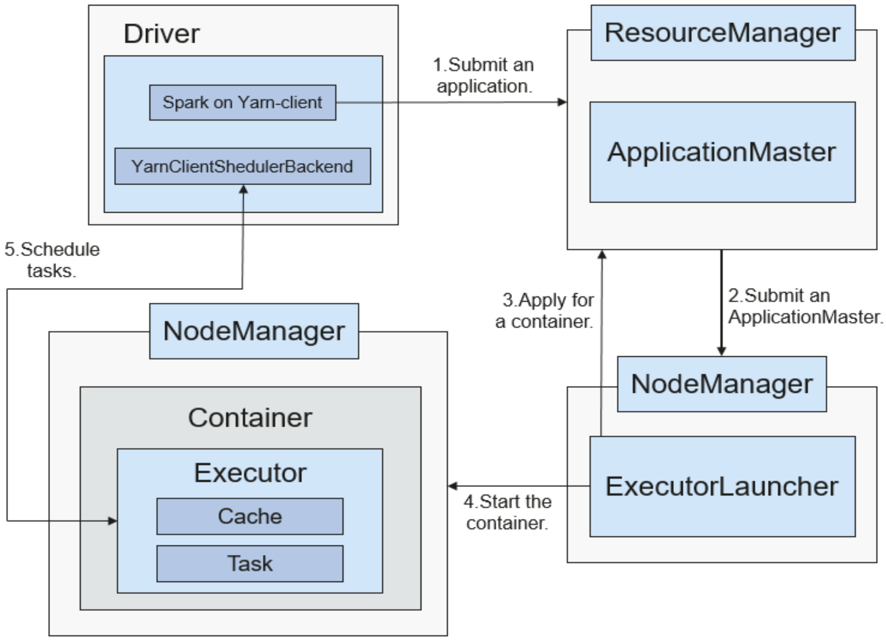
d. The driver allocates tasks to the executor.

e. The executor runs tasks and reports the operating status to the driver.

- Yarn-client mode

  Figure 1-101 shows the running framework of Spark on Yarn-cluster.

Figure 1-101 Spark on Yarn-client operation framework



Spark on Yarn-client implementation process:

NOTE

In Yarn-client mode, Driver is deployed on the client and started on the client. In Yarn-client mode, the client of the earlier version is incompatible. You are advised to use the Yarn-cluster mode.

a. The client sends the Spark application request to ResourceManager, then ResourceManager returns the results. The results include information such as Application ID and the maximum and minimum available resources. The client packages all information required to start ApplicationMaster, and sends the information to ResourceManager.

b. After receiving the request, ResourceManager finds a proper node for ApplicationMaster and starts it on this node. ApplicationMaster is a role in Yarn, and the process name in Spark is ExecutorLauncher.

c. Based on the resource requirements of each task, ApplicationMaster can apply for a series of Containers to run tasks from ResourceManager.

d. After receiving the newly allocated container list (from ResourceManager), ApplicationMaster sends information to the related NodeManagers to start the containers.

   ResourceManager allocates the containers to ApplicationMaster, which communicates with the related NodeManagers, and starts the executors in the obtained containers. After the executors are started, it registers with drivers and applies for tasks.

NOTE

Running containers are not suspended and resources are not released.

e. The drivers allocate tasks to the executors. The executor executes tasks and reports the operating status to the driver.

1.3.22.4 Spark Enhanced Open Source Feature: Optimized SQL Query of Cross-Source Data

Scenario

Enterprises usually store massive data, such as from various databases and warehouses, for management and information collection. However, diversified data sources, hybrid dataset structures, and scattered data storage lower query efficiency.

The open source Spark only supports simple filter pushdown during querying of multi-source data. The SQL engine performance is deteriorated due of a large amount of unnecessary data transmission. The pushdown function is enhanced, so that **aggregate**, complex **projection**, and complex **predicate** can be pushed to data sources, reducing unnecessary data transmission and improving query performance.

Only the JDBC data source supports pushdown of query operations, such as **aggregate**, **projection**, **predicate**, **aggregate over inner join**, and **aggregate over union all**. All pushdown operations can be enabled based on your requirements.

Table 1-20 Enhanced query of cross-source query

Module	Before Enhancement	After Enhancement
aggregate	The pushdown of **aggregate** is not supported.	● Aggregation functions including **sum**, **avg**, **max**, **min**, and **count** are supported. Example: select count(*) from table ● Internal expressions of aggregation functions are supported. Example: select sum(a+b) from table ● Calculation of aggregation functions is supported. Example: select avg(a) + max(b) from table ● Pushdown of **having** is supported. Example: select sum(a) from table where a>0 group by b having sum(a)>10 ● Pushdown of some functions is supported. Pushdown of lines in mathematics, time, and string functions, such as **abs()**, **month()**, and **length()** are supported. In addition to the preceding built-in functions, you can run the **SET** command to add functions supported by data sources. Example: select sum(abs(a)) from table ● Pushdown of **limit** and **order by** after **aggregate** is supported. However, the pushdown is not supported in Oracle, because Oracle does not support **limit**. Example: select sum(a) from table where a>0 group by b order by sum(a) limit 5
projection	Only pushdown of simple **projection** is supported. Example: select a, b from table	● Complex expressions can be pushed down. Example: select (a+b)*c from table ● Some functions can be pushed down. For details, see the description below the table. Example: select length(a)+abs(b) from table ● Pushdown of **limit** and **order by** after **projection** is supported. Example: select a, b+c from table order by a limit 3

Module	Before Enhancement	After Enhancement
predicate	Only simple filtering with the column name on the left of the operator and values on the right is supported. Example: select * from table where a>0 or b in ("aaa", "bbb")	- Complex expression pushdown is supported. Example: select * from table where a +b>c*d or a/c in (1, 2, 3) - Some functions can be pushed down. For details, see the description below the table. Example: select * from table where length(a)>5
aggregate over inner join	Related data from the two tables must be loaded to Spark. The join operation must be performed before the **aggregate** operation.	The following functions are supported: - Aggregation functions including **sum**, **avg**, **max**, **min**, and **count** are supported. - All **aggregate** operations can be performed in a same table. The **group by** operations can be performed on one or two tables and only inner join is supported. The following scenarios are not supported: - **aggregate** cannot be pushed down from both the left- and right-join tables. - **aggregate** contains operations, for example, sum(a+b). - **aggregate** operations, for example, sum(a)+min(b).
aggregate over union all	Related data from the two tables must be loaded to Spark. **union** must be performed before **aggregate**.	Supported scenarios: Aggregation functions including **sum**, **avg**, **max**, **min**, and **count** are supported. Unsupported scenarios: - **aggregate** contains operations, for example, sum(a+b). - **aggregate** operations, for example, sum(a)+min(b).

Precautions

- If external data source is Hive, query operation cannot be performed on foreign tables created by Spark.
- Only MySQL and MPPDB data sources are supported.

1.3.23 Spark2x

1.3.23.1 Basic Principles of Spark2x

NOTE

The Spark2x component applies to MRS 3.x and later versions.

Description

Spark is a memory-based distributed computing framework. In iterative computation scenarios, the computing capability of Spark is 10 to 100 times higher than MapReduce, because data is stored in memory when being processed. Spark can use HDFS as the underlying storage system, enabling users to quickly switch to Spark from MapReduce. Spark provides one-stop data analysis capabilities, such as the streaming processing in small batches, offline batch processing, SQL query, and data mining. Users can seamlessly use these functions in a same application. For details about the new open-source features of Spark2x, see **Spark2x Open Source New Features**.

Features of Spark are as follows:

- Improves the data processing capability through distributed memory computing and directed acyclic graph (DAG) execution engine. The delivered performance is 10 to 100 times higher than that of MapReduce.
- Supports multiple development languages (Scala/Java/Python) and dozens of highly abstract operators to facilitate the construction of distributed data processing applications.
- Builds data processing stacks using **SQL**, **Streaming**, MLlib, and GraphX to provide one-stop data processing capabilities.
- Fits into the Hadoop ecosystem, allowing Spark applications to run on Standalone, Mesos, or Yarn, enabling access of multiple data sources such as HDFS, HBase, and Hive, and supporting smooth migration of the MapReduce application to Spark.

Architecture

Figure 1-102 describes the Spark architecture and **Table 1-21** lists the Spark modules.

Figure 1-102 Spark architecture

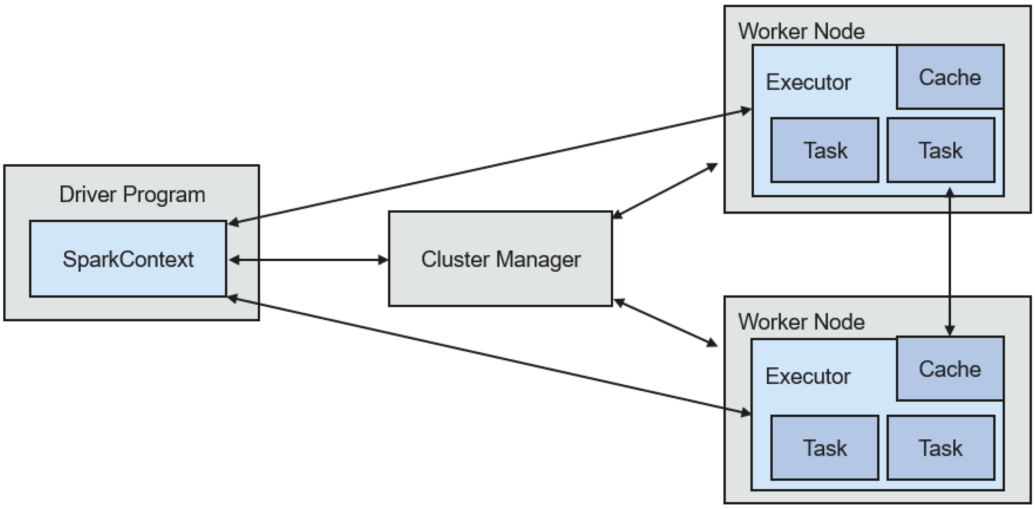


Table 1-21 Basic concepts

Module	Description
Cluster Manager	Cluster manager manages resources in the cluster. Spark supports multiple cluster managers, including Mesos, Yarn, and the Standalone cluster manager that is delivered with Spark. By default, Spark clusters adopt the Yarn cluster manager.
Application	Spark application. It consists of one Driver Program and multiple executors.
Deploy Mode	Deployment in cluster or client mode. In cluster mode, the driver runs on a node inside the cluster. In client mode, the driver runs on the client (outside the cluster).
Driver Program	The main process of the Spark application. It runs the **main()** function of an application and creates SparkContext. It is used for parsing applications, generating stages, and scheduling tasks to executors. Usually, SparkContext represents Driver Program.
Executor	A process started on a Work Node. It is used to execute tasks, and manage and process the data used in applications. A Spark application usually contains multiple executors. Each executor receives commands from the driver and executes one or multiple tasks.
Worker Node	A node that starts and manages executors and resources in a cluster.
Job	A job consists of multiple concurrent tasks. One action operator (for example, a collect operator) maps to one job.
Stage	Each job consists of multiple stages. Each stage is a task set, which is separated by Directed Acyclic Graph (DAG).

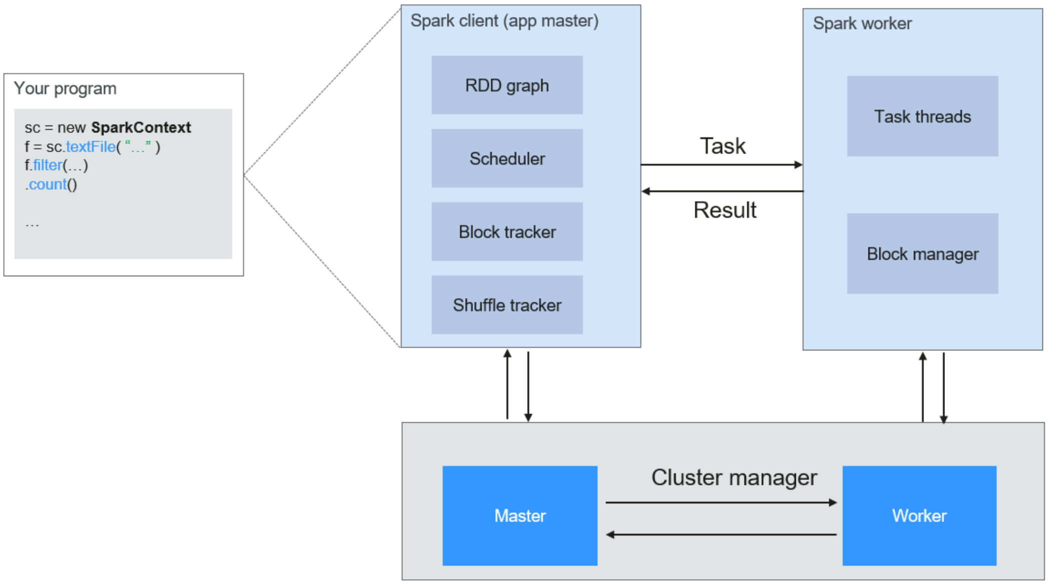
Module	Description
Task	A task carries the computation unit of the service logics. It is the minimum working unit that can be executed on the Spark platform. An application can be divided into multiple tasks based on the execution plan and computation amount.

Spark Principle

Figure 1-103 describes the application running architecture of Spark.

1. An application is running in the cluster as a collection of processes. Driver coordinates the running of the application.
2. To run an application, Driver connects to the cluster manager (such as Standalone, Mesos, and Yarn) to apply for the executor resources, and start ExecutorBackend. The cluster manager schedules resources between different applications. Driver schedules DAGs, divides stages, and generates tasks for the application at the same time.
3. Then, Spark sends the codes of the application (the codes transferred to **SparkContext**, which is defined by JAR or Python) to an executor.
4. After all tasks are finished, the running of the user application is stopped.

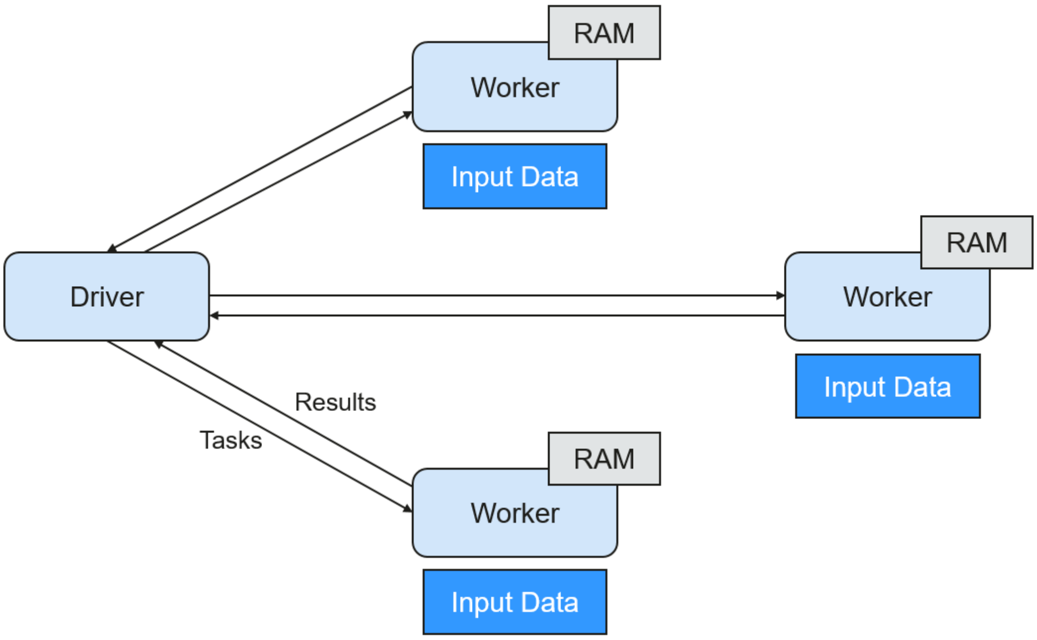
Figure 1-103 Spark application running architecture



```
sc = new SparkContext
f = sc.textFile("…")
f.filter(…)
.count()
...
```

Spark uses Master and Worker modes, as shown in **Figure 1-104**. A user submits an application on the Spark client, and then the scheduler divides a job into multiple tasks and sends the tasks to each Worker for execution. Each Worker reports the computation results to Driver (Master), and then the Driver aggregates and returns the results to the client.

Figure 1-104 Spark Master-Worker mode



Note the following about the architecture:

- Applications are isolated from each other.

  Each application has an independent executor process, and each executor starts multiple threads to execute tasks in parallel. Each driver schedules its own tasks, and different application tasks run on different JVMs, that is, different executors.
- Different Spark applications do not share data, unless data is stored in the external storage system such as HDFS.
- You are advised to deploy the Driver program in a location that is close to the Worker node because the Driver program schedules tasks in the cluster. For example, deploy the Driver program on the network where the Worker node is located.

Spark on YARN can be deployed in two modes:

- In Yarn-cluster mode, the Spark driver runs inside an ApplicationMaster process which is managed by Yarn in the cluster. After the ApplicationMaster is started, the client can exit without interrupting service running.
- In Yarn-client mode, Driver runs in the client process, and the ApplicationMaster process is used only to apply for requesting resources from Yarn.

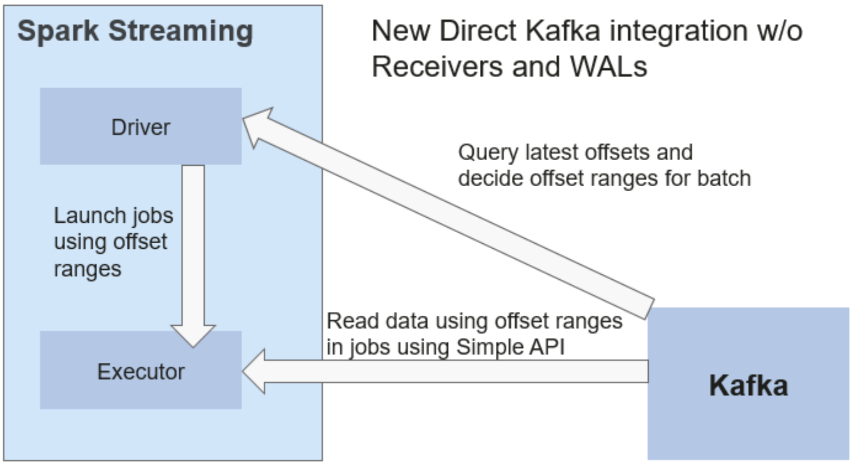
Spark Streaming Principle

Spark Streaming is a real-time computing framework built on the Spark, which expands the capability for processing massive streaming data. Spark supports two data processing approaches: Direct Streaming and Receiver.

Direct Streaming computing process

In Direct Streaming approach, Direct API is used to process data. Take Kafka Direct API as an example. Direct API provides offset location that each batch range will read from, which is much simpler than starting a receiver to continuously receive data from Kafka and written data to write-ahead logs (WALs). Then, each batch job is running and the corresponding offset data is ready in Kafka. These offset information can be securely stored in the checkpoint file and read by applications that failed to start.

Figure 1-105 Data transmission through Direct Kafka API


Spark Streaming
New Direct Kafka integration w/o Receivers and WALs
Driver
Query latest offsets and decide offset ranges for batch
Launch jobs using offset ranges
Read data using offset ranges in jobs using Simple API
Executor
Kafka

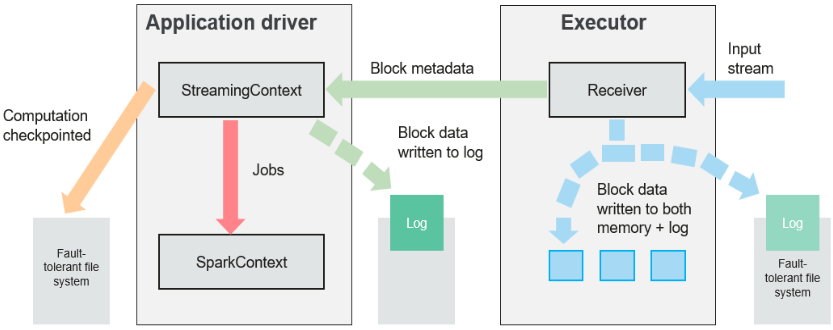
After the failure, Spark Streaming can read data from Kafka again and process the data segment. The processing result is the same no matter Spark Streaming fails or not, because the semantic is processed only once.

Direct API does not need to use the WAL and Receivers, and ensures that each Kafka record is received only once, which is more efficient. In this way, the Spark Streaming and Kafka can be well integrated, making streaming channels be featured with high fault-tolerance, high efficiency, and ease-of-use. Therefore, you are advised to use Direct Streaming to process data.

Receiver computing process

When a Spark Streaming application starts (that is, when the driver starts), the related StreamingContext (the basis of all streaming functions) uses SparkContext to start the receiver to become a long-term running task. These receivers receive and save streaming data to the Spark memory for processing. **Figure 1-106** shows the data transfer lifecycle.

Figure 1-106 Data transfer lifecycle



1. Receive data (blue arrow).

   Receiver divides a data stream into a series of blocks and stores them in the executor memory. In addition, after WAL is enabled, it writes data to the WAL of the fault-tolerant file system.

2. Notify the driver (green arrow).

   The metadata in the received block is sent to StreamingContext in the driver. The metadata includes:

   – Block reference ID used to locate the data position in the Executor memory.

   – Block data offset information in logs (if the WAL function is enabled).

3. Process data (red arrow).

   For each batch of data, StreamingContext uses block information to generate resilient distributed datasets (RDDs) and jobs. StreamingContext executes jobs by running tasks to process blocks in the executor memory.

4. Periodically set checkpoints (orange arrows).

5. For fault tolerance, StreamingContext periodically sets checkpoints and saves them to external file systems.

Fault Tolerance

Spark and its RDD allow seamless processing of failures of any Worker node in the cluster. Spark Streaming is built on top of Spark. Therefore, the Worker node of Spark Streaming also has the same fault tolerance capability. However, Spark Streaming needs to run properly in case of long-time running. Therefore, Spark must be able to recover from faults through the driver process (main process that coordinates all Workers). This poses challenges to the Spark driver fault-tolerance because the Spark driver may be any user application implemented in any computation mode. However, Spark Streaming has internal computation architecture. That is, it periodically executes the same Spark computation in each batch data. Such architecture allows it to periodically store checkpoints to reliable storage space and recover them upon the restart of Driver.

For source data such as files, the Driver recovery mechanism can ensure zero data loss because all data is stored in a fault-tolerant file system such as HDFS. However, for other data sources such as Kafka and Flume, some received data is cached only in memory and may be lost before being processed. This is caused by the distribution operation mode of Spark applications. When the driver process fails, all executors running in the Cluster Manager, together with all data in the memory, are terminated. To avoid such data loss, the WAL function is added to Spark Streaming.

WAL is often used in databases and file systems to ensure persistence of any data operation. That is, first record an operation to a persistent log and perform this operation on data. If the operation fails, the system is recovered by reading the log and re-applying the preset operation. The following describes how to use WAL to ensure persistence of received data:

Receiver is used to receive data from data sources such as Kafka. As a long-time running task in Executor, Receiver receives data, and also confirms received data if supported by data sources. Received data is stored in the Executor memory, and Driver delivers a task to Executor for processing.

After WAL is enabled, all received data is stored to log files in the fault-tolerant file system. Therefore, the received data does not lose even if Spark Streaming fails. Besides, receiver checks correctness of received data only after the data is pre-written into logs. Data that is cached but not stored can be sent again by data sources after the driver restarts. These two mechanisms ensure zero data loss. That is, all data is recovered from logs or re-sent by data sources.

To enable the WAL function, perform the following operations:

- Set **streamingContext.checkpoint** (path-to-directory) to configure the checkpoint directory, which is an HDFS file path used to store streaming checkpoints and WALs.
- Set **spark.streaming.receiver.writeAheadLog.enable** of SparkConf to **true** (the default value is **false**).

After WAL is enabled, all receivers have the advantage of recovering from reliable received data. You are advised to disable the multi-replica mechanism because the fault-tolerant file system of WAL may also replicate the data.

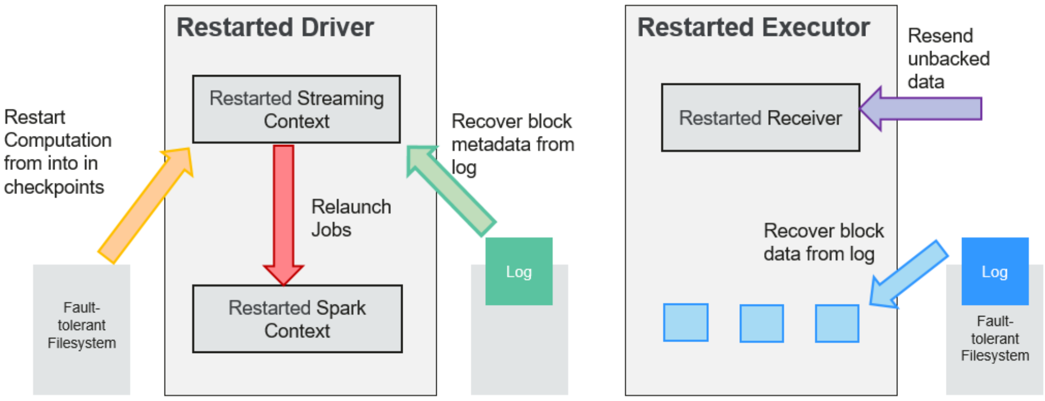
NOTE

The data receiving throughput is lowered after WAL is enabled. All data is written into the fault-tolerant file system. As a result, the write throughput of the file system and the network bandwidth for data replication may become the potential bottleneck. To solve this problem, you are advised to create more receivers to increase the degree of data receiving parallelism or use better hardware to improve the throughput of the fault-tolerant file system.

Recovery Process

When a failed driver is restarted, restart it as follows:

Figure 1-107 Computing recovery process


Restarted Driver
Restarted Streaming Context
Relaunch Jobs
Restarted Spark Context
Restart Computation from into in checkpoints
Fault-tolerant Filesystem
Recover block metadata from log
Log
Restarted Executor
Resend unbacked data
Restarted Receiver
Recover block data from log
Log
Fault-tolerant Filesystem

1. Recover computing. (Orange arrow)

   Use checkpoint information to restart Driver, reconstruct SparkContext and restart Receiver.

2. Recover metadata block. (Green arrow)

   This operation ensures that all necessary metadata blocks are recovered to continue the subsequent computing recovery.

3. Relaunch unfinished jobs. (Red arrow)

   Recovered metadata is used to generate RDDs and corresponding jobs for interrupted batch processing due to failures.

4. Read block data saved in logs. (Blue arrow)

   Block data is directly read from WALs during execution of the preceding jobs, and therefore all essential data reliably stored in logs is recovered.

5. Resend unconfirmed data. (Purple arrow)

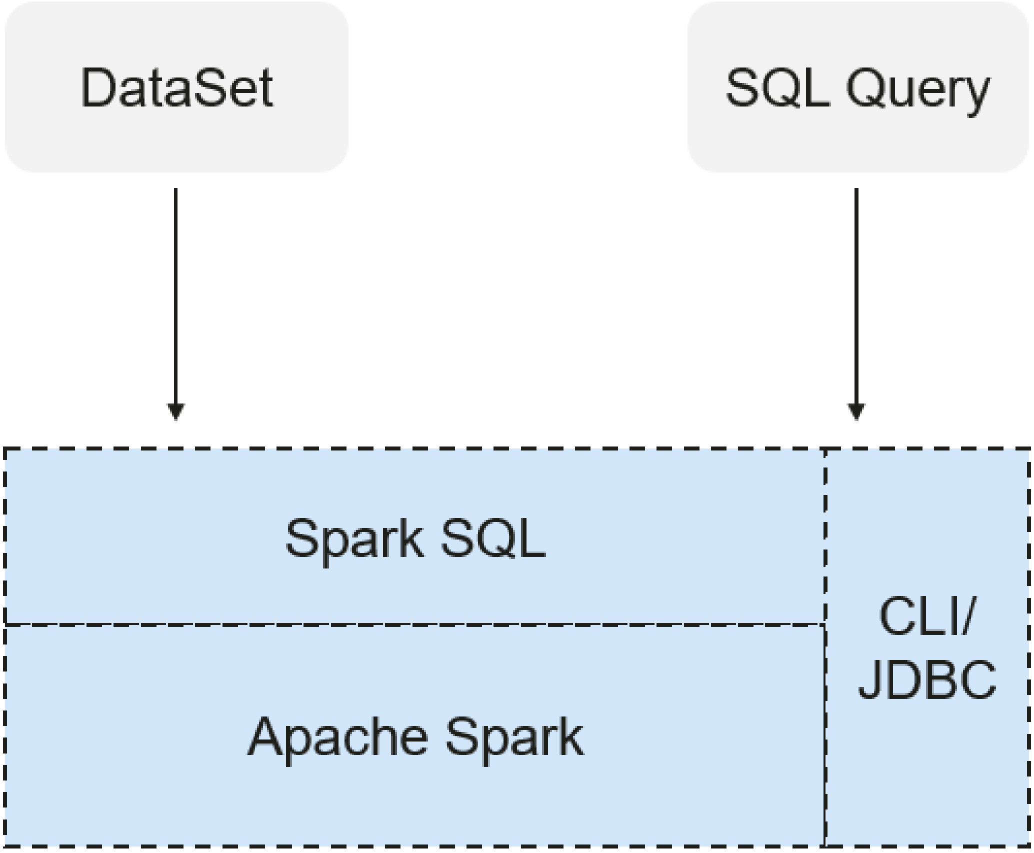
   Data that is cached but not stored to logs upon failures is re-sent by data sources, because the receiver does not confirm the data.

Therefore, by using WALs and reliable Receiver, Spark Streaming can avoid input data loss caused by Driver failures.

SparkSQL and DataSet Principle

SparkSQL

Figure 1-108 SparkSQL and DataSet



Spark SQL is a module for processing structured data. In Spark application, SQL statements or DataSet APIs can be seamlessly used for querying structured data.

Spark SQL and DataSet also provide a universal method for accessing multiple data sources such as Hive, CSV, Parquet, ORC, JSON, and JDBC. These data sources also allow data interaction. Spark SQL reuses the Hive frontend processing logic and metadata processing module. With the Spark SQL, you can directly query existing Hive data.

In addition, Spark SQL also provides API, CLI, and JDBC APIs, allowing diverse accesses to the client.

Spark SQL Native DDL/DML

In Spark 1.5, lots of Data Definition Language (DDL)/Data Manipulation Language (DML) commands are pushed down to and run on the Hive, causing coupling with the Hive and inflexibility such as unexpected error reports and results.

Spark2x realizes command localization and replaces the Hive with Spark SQL Native DDL/DML to run DDL/DML commands. Additionally, the decoupling from the Hive is realized and commands can be customized.

DataSet

A DataSet is a strongly typed collection of domain-specific objects that can be transformed in parallel using functional or relational operations. Each Dataset also has an untyped view called a DataFrame, which is a Dataset of Row.

The DataFrame is a structured and distributed dataset consisting of multiple columns. The DataFrame is equal to a table in the relationship database or the DataFrame in the R/Python. The DataFrame is the most basic concept in the Spark SQL, which can be created by using multiple methods, such as the structured dataset, Hive table, external database or RDD.

Operations available on DataSets are divided into transformations and actions.

- A transformation operation can generate a new DataSet,

  for example, **map**, **filter**, **select**, and **aggregate (groupBy)**.
- An action operation can trigger computation and return results,

  for example, **count**, **show**, or write data to the file system.

You can use either of the following methods to create a DataSet:

- The most common way is by pointing Spark to some files on storage systems, using the **read** function available on a SparkSession.
  ```
  val people = spark.read.parquet("...").as[Person]  // Scala
  DataSet<Person> people = spark.read().parquet("...").as(Encoders.bean(Person.class));//Java
  ```
- You can also create a DataSet using the transformation operation available on an existing one. For example, apply the map operation on an existing DataSet to create a DataSet:
  ```
  val names = people.map(_.name) // In Scala: names is Dataset.
  Dataset<String> names = people.map((Person p) -> p.name, Encoders.STRING)); // Java
  ```

CLI and JDBCServer

In addition to programming APIs, Spark SQL also provides the CLI/JDBC APIs.

- Both **spark-shell** and **spark-sql** scripts can provide the CLI for debugging.
- JDBCServer provides JDBC APIs. External systems can directly send JDBC requests to calculate and parse structured data.

SparkSession Principle

SparkSession is a unified API in Spark2x and can be regarded as a unified entry for reading data. SparkSession provides a single entry point to perform many operations that were previously scattered across multiple classes, and also provides accessor methods to these older classes to maximize compatibility.

A SparkSession can be created using a builder pattern. The builder will automatically reuse the existing SparkSession if there is a SparkSession; or create a SparkSession if it does not exist. During I/O transactions, the configuration item settings in the builder are automatically synchronized to Spark and Hadoop.

```
import org.apache.spark.sql.SparkSession
val sparkSession = SparkSession.builder
  .master("local")
  .appName("my-spark-app")
  .config("spark.some.config.option", "config-value")
  .getOrCreate()
```

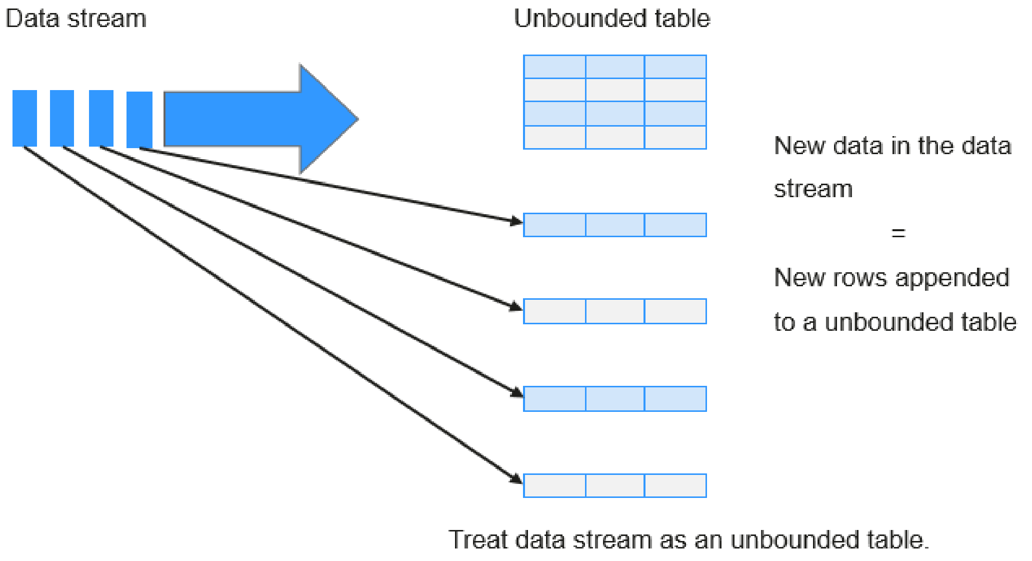
- SparkSession can be used to execute SQL queries on data and return results as DataFrame.
  ```
  sparkSession.sql("select * from person").show
  ```
- SparkSession can be used to set configuration items during running. These configuration items can be replaced with variables in SQL statements.
  ```
  sparkSession.conf.set("spark.some.config", "abcd")
  sparkSession.conf.get("spark.some.config")
  sparkSession.sql("select ${spark.some.config}")
  ```
- SparkSession also includes a "catalog" method that contains methods to work with Metastore (data catalog). After this method is used, a dataset is returned, which can be run using the same Dataset API.
  ```
  val tables = sparkSession.catalog.listTables()
  val columns = sparkSession.catalog.listColumns("myTable")
  ```
- Underlying SparkContext can be accessed by SparkContext API of SparkSession.
  ```
  val sparkContext = sparkSession.sparkContext
  ```

Structured Streaming Principle

Structured Streaming is a stream processing engine built on the Spark SQL engine. You can use the Dataset/DataFrame API in Scala, Java, Python, or R to express streaming aggregations, event-time windows, and stream-stream joins. If streaming data is incrementally and continuously produced, Spark SQL will continue to process the data and synchronize the result to the result set. In addition, the system ensures end-to-end exactly-once fault-tolerance guarantees through checkpoints and WALs.

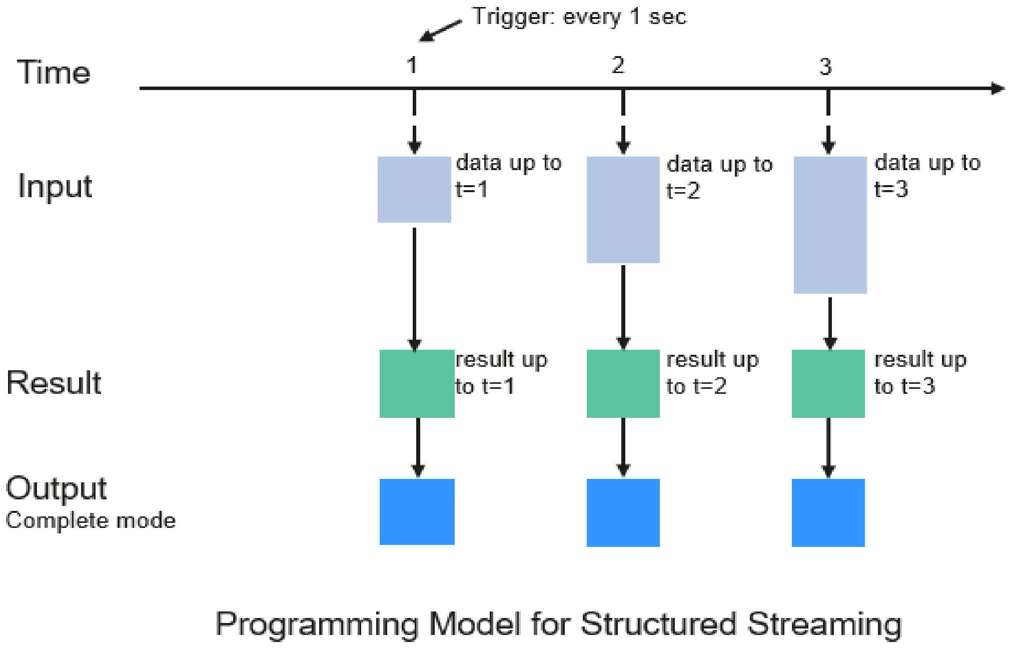
The core of Structured Streaming is to take streaming data as an incremental database table. Similar to the data block processing model, the streaming data processing model applies query operations on a static database table to streaming computing, and Spark uses standard SQL statements for query, to obtain data from the incremental and unbounded table.

Figure 1-109 Unbounded table of Structured Streaming


Data stream
Unbounded table
New data in the data stream
=
New rows appended to a unbounded table
Treat data stream as an unbounded table.

Each query operation will generate a result table. At each trigger interval, updated data will be synchronized to the result table. Whenever the result table is updated, the updated result will be written into an external storage system.

Figure 1-110 Structured Streaming data processing model



Storage modes of Structured Streaming at the output phase are as follows:

- Complete Mode: The updated result sets are written into the external storage system. The write operation is performed by a connector of the external storage system.

- Append Mode: If an interval is triggered, only added data in the result table will be written into an external system. This is applicable only on the queries where existing rows in the result table are not expected to change.
- Update Mode: If an interval is triggered, only updated data in the result table will be written into an external system, which is the difference between the Complete Mode and Update Mode.

Concepts

- **RDD**

  Resilient Distributed Dataset (RDD) is a core concept of Spark. It indicates a read-only and partitioned distributed dataset. Partial or all data of this dataset can be cached in the memory and reused between computations.

  RDD Creation

  - An RDD can be created from the input of HDFS or other storage systems that are compatible with Hadoop.
  - A new RDD can be converted from a parent RDD.
  - An RDD can be converted from a collection of datasets through encoding.

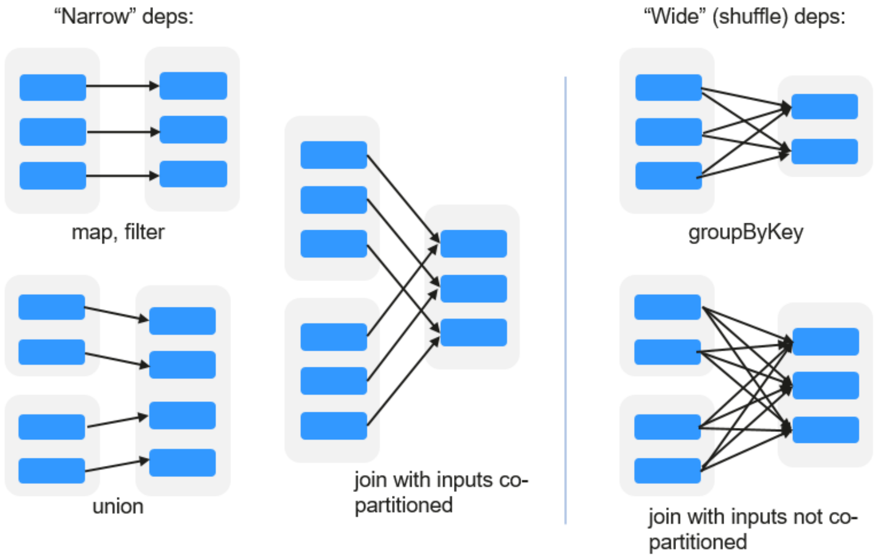
  RDD Storage

  - You can select different storage levels to store an RDD for reuse. (There are 11 storage levels to store an RDD.)
  - By default, the RDD is stored in the memory. When the memory is insufficient, the RDD overflows to the disk.

- **RDD Dependency**

  The RDD dependency includes the narrow dependency and wide dependency.

Figure 1-111 RDD dependency


"Narrow" deps:
"Wide" (shuffle) deps:
map, filter
groupByKey
union
join with inputs co-partitioned
join with inputs not co-partitioned

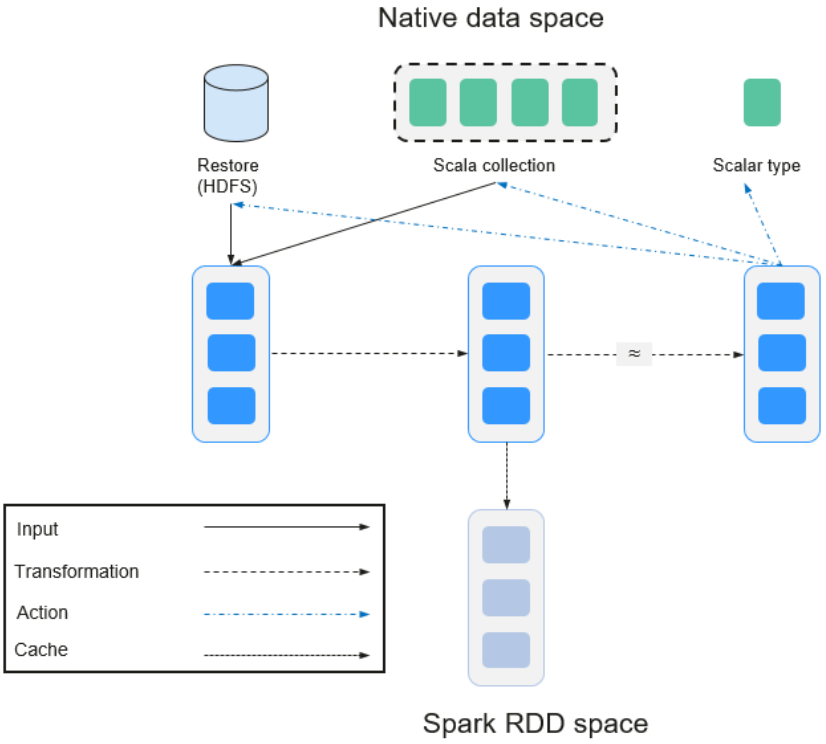
  - **Narrow dependency**: Each partition of the parent RDD is used by at most one partition of the child RDD.
  - **Wide dependency**: Partitions of the child RDD depend on all partitions of the parent RDD.

The narrow dependency facilitates the optimization. Logically, each RDD operator is a fork/join (the join is not the join operator mentioned above but the barrier used to synchronize multiple concurrent tasks); fork the RDD to each partition, and then perform the computation. After the computation, join the results, and then perform the fork/join operation on the next RDD operator. It is uneconomical to directly translate the RDD into physical implementation. The first is that every RDD (even intermediate result) needs to be physicalized into memory or storage, which is time-consuming and occupies much space. The second is that as a global barrier, the join operation is very expensive and the entire join process will be slowed down by the slowest node. If the partitions of the child RDD narrowly depend on that of the parent RDD, the two fork/join processes can be combined to implement classic fusion optimization. If the relationship in the continuous operator sequence is narrow dependency, multiple fork/join processes can be combined to reduce a large number of global barriers and eliminate the physicalization of many RDD intermediate results, which greatly improves the performance. This is called pipeline optimization in Spark.

- **Transformation and Action (RDD Operations)**

  Operations on RDD include transformation (the return value is an RDD) and action (the return value is not an RDD). **Figure 1-112** shows the RDD operation process. The transformation is lazy, which indicates that the transformation from one RDD to another RDD is not immediately executed. Spark only records the transformation but does not execute it immediately. The real computation is started only when the action is started. The action returns results or writes the RDD data into the storage system. The action is the driving force for Spark to start the computation.

Figure 1-112 RDD operation



The data and operation model of RDD are quite different from those of Scala.

```
val file = sc.textFile("hdfs://...")
val errors = file.filter(_.contains("ERROR"))
errors.cache()
errors.count()
```

a. The textFile operator reads log files from the HDFS and returns files (as an RDD).

b. The filter operator filters rows with **ERROR** and assigns them to errors (a new RDD). The filter operator is a transformation.

c. The cache operator caches errors for future use.

d. The count operator returns the number of rows of errors. The count operator is an action.

Transformation includes the following types:

– The RDD elements are regarded as simple elements.

The input and output has the one-to-one relationship, and the partition structure of the result RDD remains unchanged, for example, map.

The input and output has the one-to-many relationship, and the partition structure of the result RDD remains unchanged, for example, flatMap (one element becomes a sequence containing multiple elements after map and then flattens to multiple elements).

The input and output has the one-to-one relationship, but the partition structure of the result RDD changes, for example, union (two RDDs integrates to one RDD, and the number of partitions becomes the sum of the number of partitions of two RDDs) and coalesce (partitions are reduced).

Operators of some elements are selected from the input, such as filter, distinct (duplicate elements are deleted), subtract (elements only exist in this RDD are retained), and sample (samples are taken).

– The RDD elements are regarded as key-value pairs.

Perform the one-to-one calculation on the single RDD, such as mapValues (the partition mode of the source RDD is retained, which is different from map).

Sort the single RDD, such as sort and partitionBy (partitioning with consistency, which is important to the local optimization).

Restructure and reduce the single RDD based on key, such as groupByKey and reduceByKey.

Join and restructure two RDDs based on the key, such as join and cogroup.

📖 **NOTE**

The later three operations involving sorting are called shuffle operations.

Action includes the following types:

– Generate scalar configuration items, such as **count** (the number of elements in the returned RDD), **reduce**, **fold/aggregate** (the number of scalar configuration items that are returned), and **take** (the number of elements before the return).

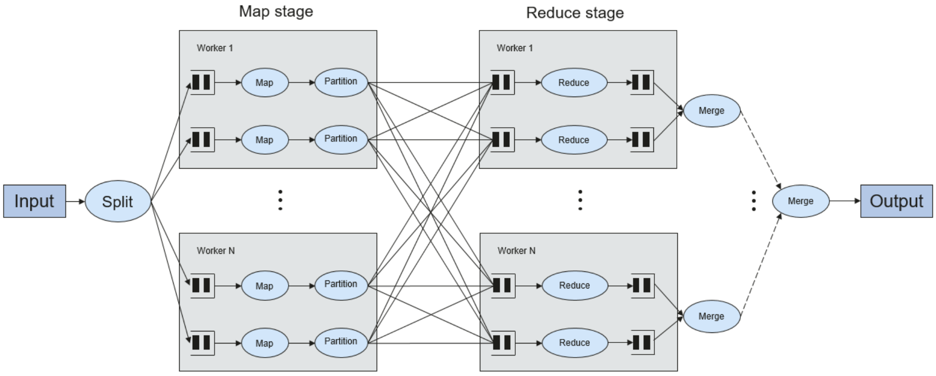
- Generate the Scala collection, such as **collect** (import all elements in the RDD to the Scala collection) and **lookup** (look up all values corresponds to the key).
- Write data to the storage, such as **saveAsTextFile** (which corresponds to the preceding **textFile**).
- Check points, such as the **checkpoint** operator. When Lineage is quite long (which occurs frequently in graphics computation), it takes a long period of time to execute the whole sequence again when a fault occurs. In this case, checkpoint is used as the check point to write the current data to stable storage.

- **Shuffle**

  Shuffle is a specific phase in the MapReduce framework, which is located between the Map phase and the Reduce phase. If the output results of Map are to be used by Reduce, the output results must be hashed based on a key and distributed to each Reducer. This process is called Shuffle. Shuffle involves the read and write of the disk and the transmission of the network, so that the performance of Shuffle directly affects the operation efficiency of the entire program.

  The figure below shows the entire process of the MapReduce algorithm.

  Figure 1-113 Algorithm process



  Shuffle is a bridge to connect data. The following describes the implementation of shuffle in Spark.

  Shuffle divides a job of Spark into multiple stages. The former stages contain one or more ShuffleMapTasks, and the last stage contains one or more ResultTasks.

- **Spark Application Structure**

  The Spark application structure includes the initialized SparkContext and the main program.

  - Initialized SparkContext: constructs the operating environment of the Spark Application.

    Constructs the SparkContext object. The following is an example:

    ```
    new SparkContext(master, appName, [SparkHome], [jars])
    ```

    Parameter description:

    master: indicates the link string. The link modes include local, Yarn-cluster, and Yarn-client.

appName: indicates the application name.

SparkHome: indicates the directory where Spark is installed in the cluster.

jars: indicates the code and dependency package of an application.

– Main program: processes data.

For details about how to submit an application, visit **https://spark.apache.org/docs/3.1.1/submitting-applications.html**.

- **Spark Shell Commands**

  The basic Spark shell commands support the submission of Spark applications. The Spark shell commands are as follows:

  ```
  ./bin/spark-submit \
    --class <main-class> \
    --master <master-url> \
    ... # other options
    <application-jar> \
    [application-arguments]
  ```

  Parameter description:

  --class: indicates the name of the class of a Spark application.

  --master: indicates the master to which the Spark application links, such as Yarn-client and Yarn-cluster.

  application-jar: indicates the path of the JAR file of the Spark application.

  application-arguments: indicates the parameter required to submit the Spark application. This parameter can be left blank.

- **Spark JobHistory Server**

  The Spark web UI is used to monitor the details in each phase of the Spark framework of a running or historical Spark job and provide the log display, which helps users to develop, configure, and optimize the job in more fine-grained units.

1.3.23.2 Spark2x HA Solution

1.3.23.2.1 Spark2x Multi-active Instance

Background

Based on existing JDBCServers in the community, multi-active-instance HA is used to achieve the high availability. In this mode, multiple JDBCServers coexist in the cluster and the client can randomly connect any JDBCServer to perform service operations. When one or multiple JDBCServers stop working, a client can connect to another normal JDBCServer.

Compared with active/standby HA, multi-active instance HA eliminates the following restrictions:

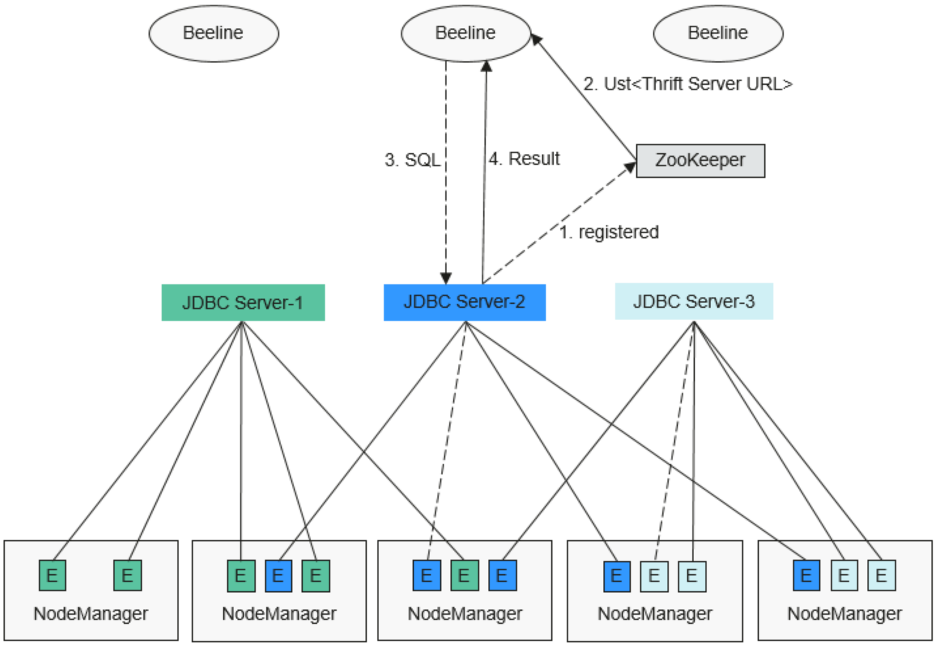
- In active/standby HA, when the active/standby switchover occurs, the unavailable period cannot be controlled by JDBCServer, but determined by Yarn service resources.
- In Spark, the Thrift JDBC similar to HiveServer2 provides services and users access services through Beeline and JDBC API. Therefore, the processing capability of the JDBCServer cluster depends on the single-point capability of the primary server, and the scalability is insufficient.

Multi-active instance HA not only prevents service interruption caused by switchover, but also enables cluster scale-out to secure high concurrency.

Implementation

The following figure shows the basic principle of multi-active instance HA of Spark JDBCServer.

Figure 1-114 Spark JDBCServer HA



1. After JDBCServer is started, it registers with ZooKeeper by writing node information in a specified directory. Node information includes the JDBCServer instance IP, port number, version, and serial number (information of different nodes is separated by commas).

   An example is provided as follows:

   ```
   [serverUri=192.168.169.84:22550
   ;version=8.1.0.1;sequence=0000001244,serverUri=192.168.195.232:22550 ;version=8.1.0.1;sequence=000
   0001242,serverUri=192.168.81.37:22550 ;version=8.1.0.1;sequence=0000001243]
   ```

2. To connect to JDBCServer, the client must specify the namespace, which is the directory of JDBCServer instances in ZooKeeper. During the connection, a JDBCServer instance is randomly selected from the specified namespace. For details about URL, see **URL Connection**.
3. After the connection succeeds, the client sends SQL statements to JDBCServer.
4. JDBCServer executes received SQL statements and sends results back to the client.

In multi-active instance HA mode, all JDBCServer instances are independent and equivalent. When one instance is interrupted during upgrade, other JDBCServer instances can accept the connection request from the client.

Following rules must be followed in the multi-active instance HA of Spark JDBCServer:

- If a JDBCServer instance exits abnormally, no other instance will take over the sessions and services running on this abnormal instance.
- When the JDBCServer process is stopped, corresponding nodes are deleted from ZooKeeper.
- The client randomly selects the server, which may result in uneven session allocation, and finally result in imbalance of instance load.
- After the instance enters the maintenance mode (in which no new connection request from the client is accepted), services still running on the instance may fail when the decommissioning times out.

URL Connection

Multi-active instance mode

In multi-active instance mode, the client reads content from the ZooKeeper node and connects to JDBCServer. The connection strings are as follows:

- Security mode:
  - If Kinit authentication is enabled, the JDBCURL is as follows:

    ```
    jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark2x/hadoop.<System domain name>@<System domain name>;
    ```

    NOTE

    - **<zkNode_IP>:<zkNode_Port>** indicates the ZooKeeper URL. Use commas (,) to separate multiple URLs,

      For example, **192.168.81.37:2181,192.168.195.232:2181,192.168.169.84:2181**.
    - **sparkthriftserver2x** indicates the directory in ZooKeeper, where a random JDBCServer instance is connected to the client.

    For example, when you use Beeline client for connection in security mode, run the following command:

    ```
    sh CLIENT_HOME/spark/bin/beeline -u "jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark2x/hadoop.<System domain name>@<System domain name>;"
    ```

  - If Keytab authentication is enabled, the JDBCURL is as follows:

    ```
    jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark2x/hadoop.<System domain name>@<System domain name>;user.principal=<principal_name>;user.keytab=<path_to_keytab>
    ```

    <principal_name> indicates the principal of Kerberos user, for example, **test@***<System domain name>*. *<path_to_keytab>* indicates the Keytab file path corresponding to *<principal_name>*, for example, **/opt/auth/test/user.keytab**.
- Common mode:

  ```
  jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;
  ```

For example, when you use Beeline client for connection in common mode, run the following command:

sh *CLIENT_HOME***/spark/bin/beeline -u "jdbc:hive2://***<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>***/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;"**

Non-multi-active instance mode

In non-multi-active instance mode, a client connects to a specified JDBCServer node. Compared with multi-active instance mode, the connection string in non-multi-active instance mode does not contain **serviceDiscoveryMode** and **zooKeeperNamespace** parameters about ZooKeeper.

For example, when you use Beeline client to connect JDBCServer in non-multi-active instance mode, run the following command:

sh *CLIENT_HOME***/spark/bin/beeline -u "jdbc:hive2://***<server_IP>:<server_Port>***/;user.principal=spark2x/hadoop.***<System domain name>@<System domain name>***;saslQop=auth-conf;auth=KERBEROS;principal=spark2x/hadoop.***<System domain name>@<System domain name>***;"**

NOTE

- **<server_IP>:<server_Port>** indicates the URL of the specified JDBCServer node.
- **CLIENT_HOME** indicates the client path.

Except the connection method, operations of JDBCServer API in multi-active instance mode and non-multi-active instance mode are the same. Spark JDBCServer is another implementation of HiveServer2 in Hive. For details about other operations, see official website of Hive at **https://cwiki.apache.org/confluence/display/Hive/HiveServer2+Clients**.

1.3.23.2.2 Spark2x Multi-tenant

Background

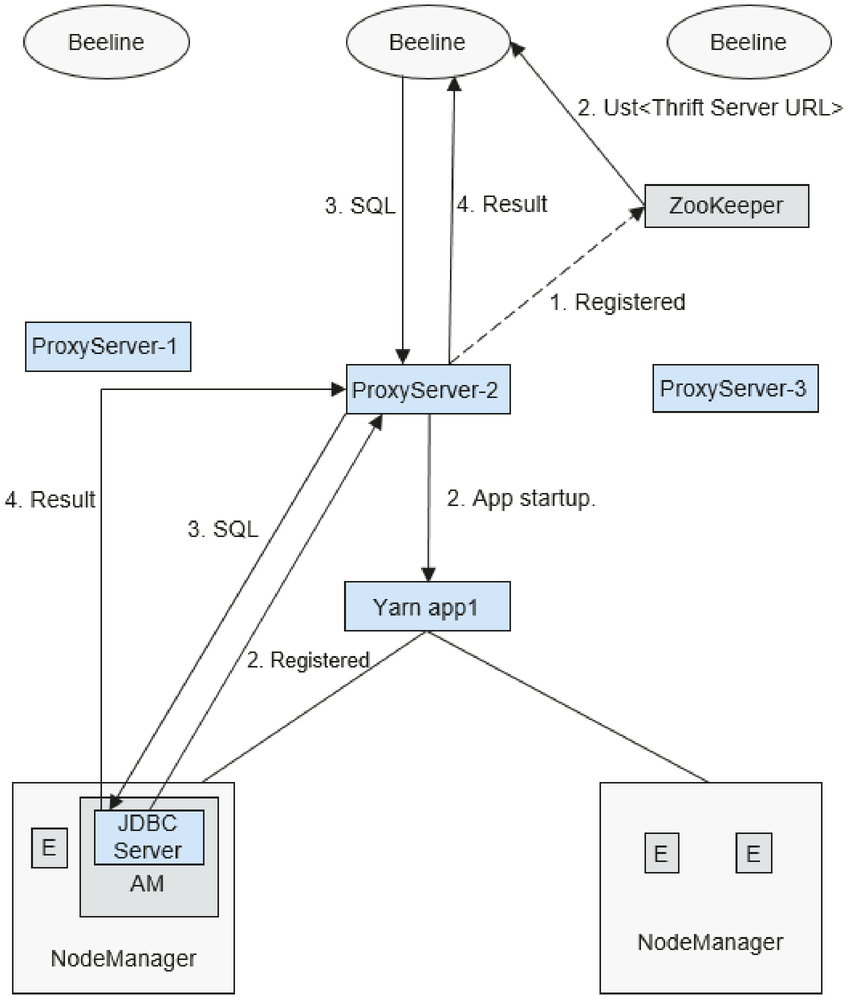
In the JDBCServer multi-active instance mode, JDBCServer implements the Yarn-client mode but only one Yarn resource queue is available. To solve the resource limitation problem, the multi-tenant mode is introduced.

In multi-tenant mode, JDBCServers are bound with tenants. Each tenant corresponds to one or more JDBCServers, and a JDBCServer provides services for only one tenant. Different tenants can be configured with different Yarn queues to implement resource isolation. In addition, JDBCServer can be dynamically started as required to avoid resource waste.

Implementation

Figure 1-115 shows the HA solution of the multi-tenant mode.

Figure 1-115 Multi-tenant mode of Spark JDBCServer



1. When ProxyServer is started, it registers with ZooKeeper by writing node information in a specified directory. Node information includes the instance IP, port number, version, and serial number (information of different nodes is separated by commas).

   NOTE

   In multi-tenant mode, the JDBCServer instance on MRS page indicates ProxyServer, the JDBCServer agent.

   An example is provided as follows:
   ```
   serverUri=192.168.169.84:22550
   ;version=8.1.0.1;sequence=0000001244,serverUri=192.168.195.232:22550
   ;version=8.1.0.1;sequence=0000001242,serverUri=192.168.81.37:22550
   ;version=8.1.0.1;sequence=0000001243,
   ```
2. To connect to ProxyServer, the client must specify a namespace, which is the directory of the ProxyServer instance that you want to access in ZooKeeper. When the client connects to ProxyServer, an instance under Namespace is randomly selected for connection. For details about the URL, see **URL Connection**.
3. After the client successfully connects to ProxyServer, ProxyServer checks whether the JDBCServer of a tenant exists. If yes, Beeline connects the

JDBCServer. If no, a new JDBCServer is started in Yarn-cluster mode. After the startup of JDBCServer, ProxyServer obtains the IP address of the JDBCServer and establishes the connection between Beeline and JDBCServer.

4. The client sends SQL statements to ProxyServer, which then forwards statements to the connected JDBCServer. JDBCServer returns the results to ProxyServer, which then returns the results to the client.

In multi-tenant HA mode, all ProxyServer instances are independent and equivalent. If one instance is interrupted during upgrade, other instances can accept the connection request from the client.

URL Connection

Multi-tenant mode

In multi-tenant mode, the client reads content from the ZooKeeper node and connects to ProxyServer. The connection strings are as follows:

- Security mode:
  - If Kinit authentication is enabled, the client URL is as follows:

    ```
    jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark2x/hadoop.<System domain name>@<System domain name>;
    ```

    NOTE

    - **<zkNode_IP>:<zkNode_Port>** indicates the ZooKeeper URL. Use commas (,) to separate multiple URLs,

      For example, **192.168.81.37:2181,192.168.195.232:2181,192.168.169.84:2181**.
    - **sparkthriftserver2x** indicates the ZooKeeper directory, where a random JDBCServer instance is connected to the client.

    For example, when you use Beeline client for connection in security mode, run the following command:

    ```
    sh CLIENT_HOME/spark/bin/beeline -u "jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark2x/hadoop.<System domain name>@<System domain name>;"
    ```
  - If Keytab authentication is enabled, the URL is as follows:

    ```
    jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark2x/hadoop.<System domain name>@<System domain name>;user.principal=<principal_name>;user.keytab=<path_to_keytab>
    ```

    <principal_name> indicates the principal of Kerberos user, for example, **test@***<System domain name>*. *<path_to_keytab>* indicates the Keytab file path corresponding to *<principal_name>*, for example, **/opt/auth/test/user.keytab**.
- Common mode:

  ```
  jdbc:hive2://<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;
  ```

For example, when you use Beeline client for connection in common mode, run the following command:

sh *CLIENT_HOME***/spark/bin/beeline -u "jdbc:hive2://** *<zkNode1_IP>:<zkNode1_Port>,<zkNode2_IP>:<zkNode2_Port>,<zkNode3_IP>:<zkNode3_Port>***/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;"**

Non-multi-tenant mode

In non-multi-tenant mode, a client connects to a specified JDBCServer node. Compared with multi-active instance mode, the connection string in non-multi-active instance mode does not contain **serviceDiscoveryMode** and **zooKeeperNamespace** parameters about ZooKeeper.

For example, when you use Beeline client to connect JDBCServer in non-multi-tenant instance mode, run the following command:

sh *CLIENT_HOME***/spark/bin/beeline -u "jdbc:hive2://** *<server_IP>:<server_Port>***/;user.principal=spark/hadoop.***<System domain name>@<System domain name>***;saslQop=auth-conf;auth=KERBEROS;principal=spark/hadoop.***<System domain name>@<System domain name>***;"**

NOTE

- **<server_IP>:<server_Port>** indicates the URL of the specified JDBCServer node.
- **CLIENT_HOME** indicates the client path.

Except the connection method, other operations of JDBCServer API in multi-tenant mode and non-multi-tenant mode are the same. Spark JDBCServer is another implementation of HiveServer2 in Hive. For details about other operations, see official website of Hive at **https://cwiki.apache.org/confluence/display/Hive/HiveServer2+Clients**.

Specifying a Tenant

Generally, the client submitted by a user connects to the default JDBCServer of the tenant to which the user belongs. If you want to connect the client to the JDBCServer of a specified tenant, add the **--hiveconf mapreduce.job.queuename** parameter.

Command for connecting Beeline is as follows (**aaa** indicates the tenant name):

beeline --hiveconf mapreduce.job.queuename=aaa -u 'jdbc:hive2://192.168.39.30:2181,192.168.40.210:2181,192.168.215.97:2181;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=sparkthriftserver2x;saslQop=auth-conf;auth=KERBEROS;principal=spark2x/hadoop.*<System domain name>*@*<System domain name>*;'

1.3.23.3 Relationship Between Spark2x and Other Components

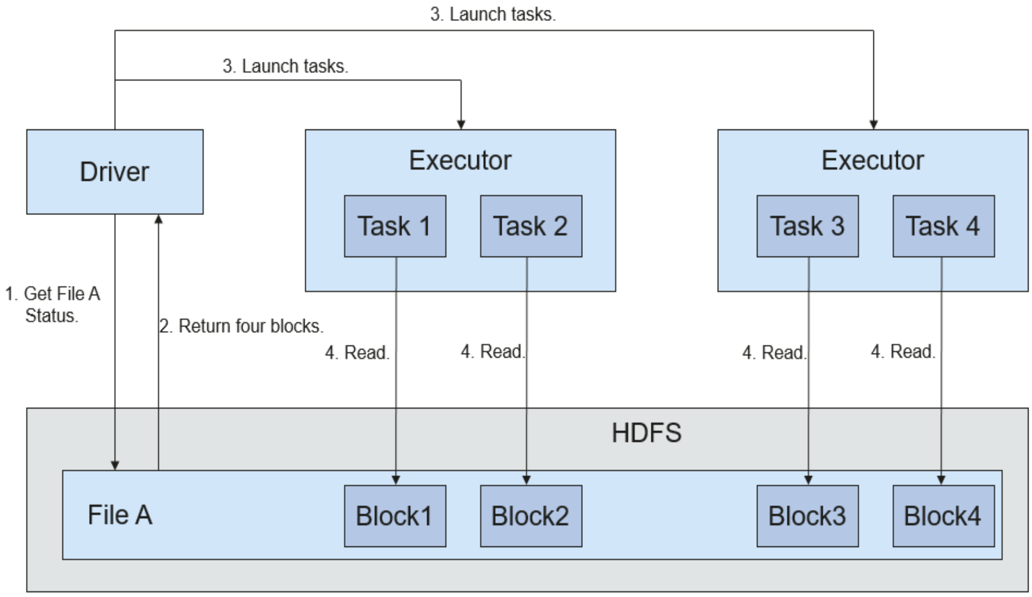
Relationship Between Spark and HDFS

Data computed by Spark comes from multiple data sources, such as local files and HDFS. Most data comes from HDFS which can read data in large scale for parallel computing After being computed, data can be stored in HDFS.

Spark involves Driver and Executor. Driver schedules tasks and Executor runs tasks.

Figure 1-116 describes the file reading process.

Figure 1-116 File reading process

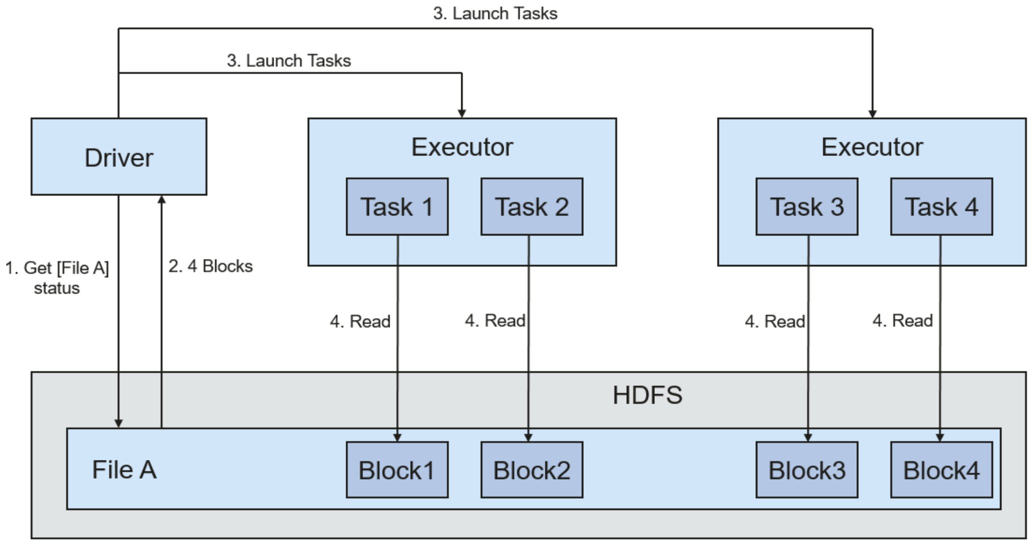


The file reading process is as follows:

1. Driver interconnects with HDFS to obtain the information of File A.
2. The HDFS returns the detailed block information about this file.
3. Driver sets a parallel degree based on the block data amount, and creates multiple tasks to read the blocks of this file.
4. Executor runs the tasks and reads the detailed blocks as part of the Resilient Distributed Dataset (RDD).

Figure 1-117 describes the file writing process.

Figure 1-117 File writing process



The file writing process is as follows:

1. Driver creates a directory where the file is to be written.
2. Based on the RDD distribution status, the number of tasks related to data writing is computed, and these tasks are sent to Executor.
3. Executor runs these tasks, and writes the RDD data to the directory created in 1.

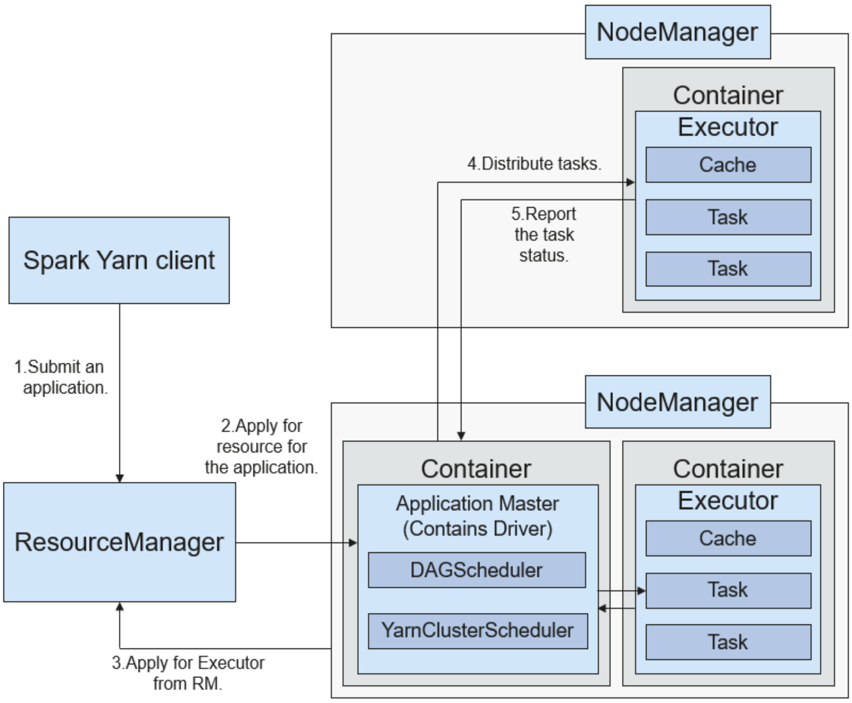
Relationship with Yarn

The Spark computing and scheduling can be implemented using Yarn mode. Spark enjoys the computing resources provided by Yarn clusters and runs tasks in a distributed way. Spark on Yarn has two modes: Yarn-cluster and Yarn-client.

- Yarn-cluster mode

  Figure 1-118 describes the operation framework.

  Figure 1-118 Spark on Yarn-cluster operation framework



  Spark on Yarn-cluster implementation process:

  a. The client generates the application information, and then sends the information to ResourceManager.
  b. ResourceManager allocates the first container (ApplicationMaster) to SparkApplication and starts the driver on the container.
  c. ApplicationMaster applies for resources from ResourceManager to run the container.

     ResourceManager allocates the containers to ApplicationMaster, which communicates with the related NodeManagers and starts the executor in the obtained container. After the executor is started, it registers with drivers and applies for tasks.

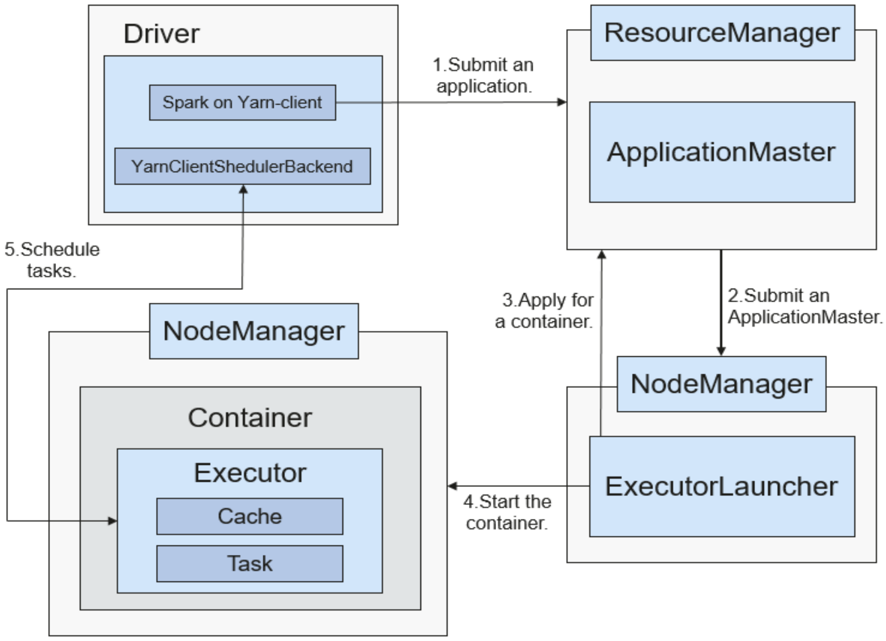
d. Drivers allocate tasks to the executors.

e. Executors run tasks and report the operating status to Drivers.

- Yarn-client mode

  Figure 1-119 describes the operation framework.

Figure 1-119 Spark on Yarn-client operation framework



Spark on Yarn-client implementation process:

NOTE

In Yarn-client mode, the Driver is deployed and started on the client. In Yarn-client mode, the client of an earlier version is incompatible. The Yarn-cluster mode is recommended.

a. The client sends the Spark application request to ResourceManager, and packages all information required to start ApplicationMaster and sends the information to ResourceManager. ResourceManager then returns the results to the client. The results include information such as ApplicationId, and the upper limit as well as lower limit of available resources. After receiving the request, ResourceManager finds a proper node for ApplicationMaster and starts it on this node. ApplicationMaster is a role in Yarn, and the process name in Spark is ExecutorLauncher.

b. Based on the resource requirements of each task, ApplicationMaster can apply for a series of containers to run tasks from ResourceManager.

c. After receiving the newly allocated container list (from ResourceManager), ApplicationMaster sends information to the related NodeManagers to start the containers.

   ResourceManager allocates the containers to ApplicationMaster, which communicates with the related NodeManagers and starts the executor in the obtained container. After the executor is started, it registers with drivers and applies for tasks.

NOTE

Running Containers will not be suspended to release resources.

d. Drivers allocate tasks to the executors. Executors run tasks and report the operating status to Drivers.

1.3.23.4 Spark2x Open Source New Features

Purpose

Compared with Spark 1.5, Spark2*x* has some new open-source features. The specific features or concepts are as follows:

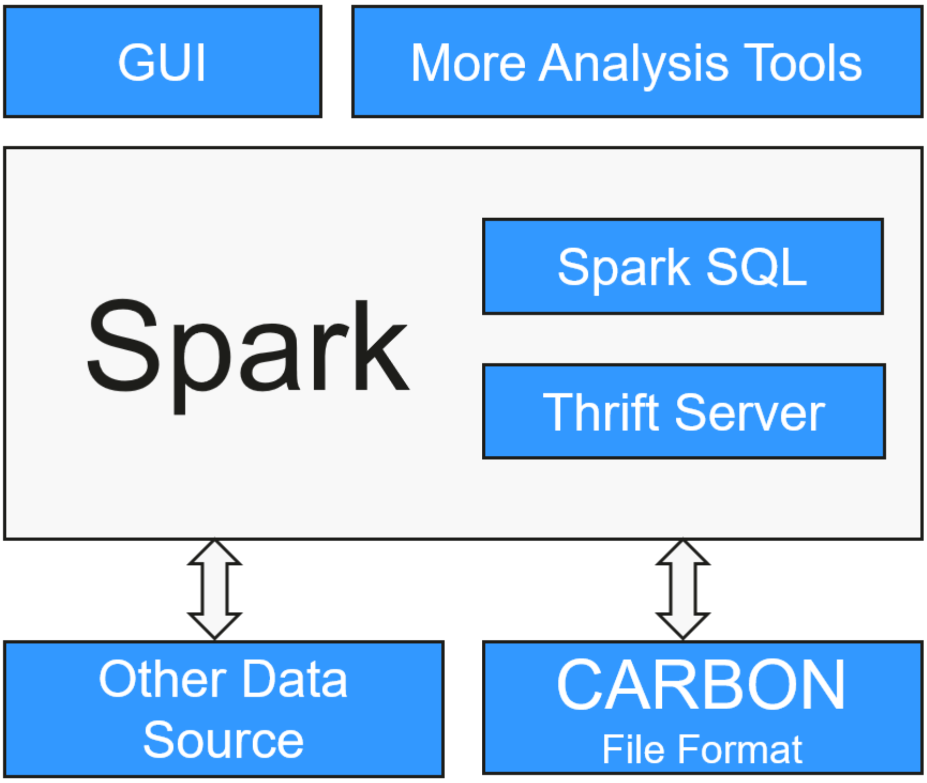
- DataSet: For details, see **SparkSQL and DataSet Principle**.
- Spark SQL Native DDL/DML: For details, see **SparkSQL and DataSet Principle**.
- SparkSession: For details, see **SparkSession Principle**.
- Structured Streaming: For details, see **Structured Streaming Principle**.
- Optimizing Small Files
- Optimizing the Aggregate Algorithm
- Optimizing Datasource Tables
- Merging CBO

1.3.23.5 Spark2x Enhanced Open Source Features

1.3.23.5.1 CarbonData Overview

CarbonData is a new Apache Hadoop native data-store format. CarbonData allows faster interactive queries over PetaBytes of data using advanced columnar storage, index, compression, and encoding techniques to improve computing efficiency. In addition, CarbonData is also a high-performance analysis engine that integrates data sources with Spark.

Figure 1-120 Basic architecture of CarbonData



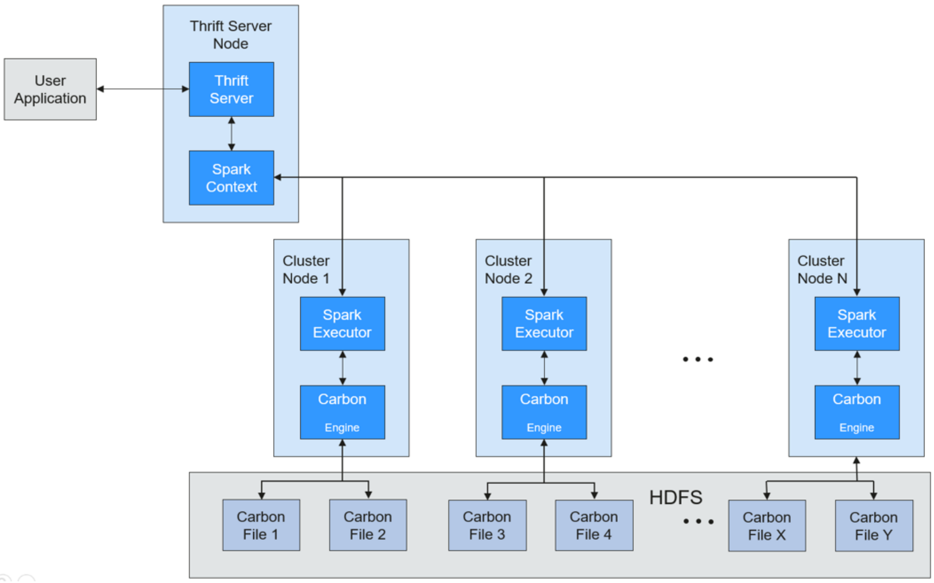
The purpose of using CarbonData is to provide quick response to ad hoc queries of big data. Essentially, CarbonData is an Online Analytical Processing (OLAP) engine, which stores data by using tables similar to those in Relational Database Management System (RDBMS). You can import more than 10 TB data to tables created in CarbonData format, and CarbonData automatically organizes and stores data using the compressed multi-dimensional indexes. After data is loaded to CarbonData, CarbonData responds to ad hoc queries in seconds.

CarbonData integrates data sources into the Spark ecosystem and you can query and analyze the data using Spark SQL. You can also use the third-party tool JDBCServer provided by Spark to connect to SparkSQL.

Topology of CarbonData

CarbonData runs as a data source inside Spark. Therefore, CarbonData does not start any additional processes on nodes in clusters. CarbonData engine runs inside the Spark executor.

Figure 1-121 Topology of CarbonData



Data stored in CarbonData Table is divided into several CarbonData data files. Each time when data is queried, CarbonData Engine reads and filters data sets. CarbonData Engine runs as a part of the Spark Executor process and is responsible for handling a subset of data file blocks.

Table data is stored in HDFS. Nodes in the same Spark cluster can be used as HDFS data nodes.

CarbonData Features

- SQL: CarbonData is compatible with Spark SQL and supports SQL query operations performed on Spark SQL.
- Simple Table dataset definition: CarbonData allows you to define and create datasets by using user-friendly Data Definition Language (DDL) statements. CarbonData DDL is flexible and easy to use, and can define complex tables.
- Easy data management: CarbonData provides various data management functions for data loading and maintenance. CarbonData supports bulk loading of historical data and incremental loading of new data. Loaded data can be deleted based on load time and a specific loading operation can be undone.
- CarbonData file format is a columnar store in HDFS. This format has many new column-based file storage features, such as table splitting and data compression. CarbonData has the following characteristics:
  - Stores data along with index: Significantly accelerates query performance and reduces the I/O scans and CPU resources, when there are filters in the query. CarbonData index consists of multiple levels of indices. A processing framework can leverage this index to reduce the task that needs to be schedules and processed, and it can also perform skip scan in more finer grain unit (called blocklet) in task side scanning instead of scanning the whole file.
  - Operable encoded data: Through supporting efficient compression and global encoding schemes, CarbonData can query on compressed/encoded

  data. The data can be converted just before returning the results to the users, which is called late materialized.
  - Supports various use cases with one single data format: like interactive OLAP-style query, sequential access (big scan), and random access (narrow scan).

Key Technologies and Advantages of CarbonData

- Quick query response: CarbonData features high-performance query. The query speed of CarbonData is 10 times of that of Spark SQL. It uses dedicated data formats and applies multiple index technologies, global dictionary code, and multiple push-down optimizations, providing quick response to TB-level data queries.
- Efficient data compression: CarbonData compresses data by combining the lightweight and heavyweight compression algorithms. This significantly saves 60% to 80% data storage space and the hardware storage cost.

CarbonData Index Cache Server

To solve the pressure and problems brought by the increasing data volume to the driver, an independent index cache server is introduced to separate the index from the Spark application side of Carbon query. All index content is managed by the index cache server. Spark applications obtain required index data in RPC mode. In this way, a large amount of memory on the service side is released so that services are not affected by the cluster scale and the performance or functions are not affected.

1.3.23.5.2 Optimizing SQL Query of Data of Multiple Sources

Scenario

Enterprises usually store massive data, such as from various databases and warehouses, for management and information collection. However, diversified data sources, hybrid dataset structures, and scattered data storage lower query efficiency.

The open source Spark only supports simple filter pushdown during querying of multi-source data. The SQL engine performance is deteriorated due of a large amount of unnecessary data transmission. The pushdown function is enhanced, so that **aggregate**, complex **projection**, and complex **predicate** can be pushed to data sources, reducing unnecessary data transmission and improving query performance.

Only the JDBC data source supports pushdown of query operations, such as **aggregate**, **projection**, **predicate**, **aggregate over inner join**, and **aggregate over union all**. All pushdown operations can be enabled based on your requirements.

Table 1-22 Enhanced query of cross-source query

Module	Before Enhancement	After Enhancement
aggregate	The pushdown of **aggregate** is not supported.	● Aggregation functions including **sum**, **avg**, **max**, **min**, and **count** are supported. Example: select count(*) from table ● Internal expressions of aggregation functions are supported. Example: select sum(a+b) from table ● Calculation of aggregation functions is supported. Example: select avg(a) + max(b) from table ● Pushdown of **having** is supported. Example: select sum(a) from table where a>0 group by b having sum(a)>10 ● Pushdown of some functions is supported. Pushdown of lines in mathematics, time, and string functions, such as **abs()**, **month()**, and **length()** are supported. In addition to the preceding built-in functions, you can run the **SET** command to add functions supported by data sources. Example: select sum(abs(a)) from table ● Pushdown of **limit** and **order by** after **aggregate** is supported. However, the pushdown is not supported in Oracle, because Oracle does not support **limit**. Example: select sum(a) from table where a>0 group by b order by sum(a) limit 5
projection	Only pushdown of simple **projection** is supported. Example: select a, b from table	● Complex expressions can be pushed down. Example: select (a+b)*c from table ● Some functions can be pushed down. For details, see the description below the table. Example: select length(a)+abs(b) from table ● Pushdown of **limit** and **order by** after **projection** is supported. Example: select a, b+c from table order by a limit 3

Module	Before Enhancement	After Enhancement
predicate	Only simple filtering with the column name on the left of the operator and values on the right is supported. Example: select * from table where a>0 or b in ("aaa", "bbb")	● Complex expression pushdown is supported. Example: select * from table where a +b>c*d or a/c in (1, 2, 3) ● Some functions can be pushed down. For details, see the description below the table. Example: select * from table where length(a)>5
aggregate over inner join	Related data from the two tables must be loaded to Spark. The join operation must be performed before the **aggregate** operation.	The following functions are supported: ● Aggregation functions including **sum**, **avg**, **max**, **min**, and **count** are supported. ● All **aggregate** operations can be performed in a same table. The **group by** operations can be performed on one or two tables and only inner join is supported. The following scenarios are not supported: ● **aggregate** cannot be pushed down from both the left- and right-join tables. ● **aggregate** contains operations, for example, sum(a+b). ● **aggregate** operations, for example, sum(a)+min(b).
aggregate over union all	Related data from the two tables must be loaded to Spark. **union** must be performed before **aggregate**.	Supported scenarios: Aggregation functions including **sum**, **avg**, **max**, **min**, and **count** are supported. Unsupported scenarios: ● **aggregate** contains operations, for example, sum(a+b). ● **aggregate** operations, for example, sum(a)+min(b).

Precautions

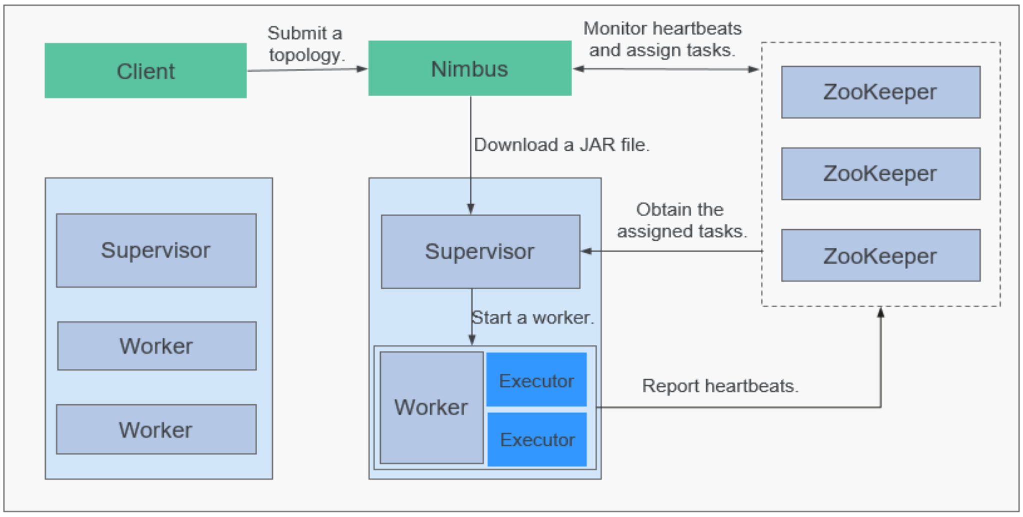
- If external data source is Hive, query operation cannot be performed on foreign tables created by Spark.
- Only MySQL and MPPDB data sources are supported.

1.3.24 Storm

1.3.24.1 Storm Basic Principles

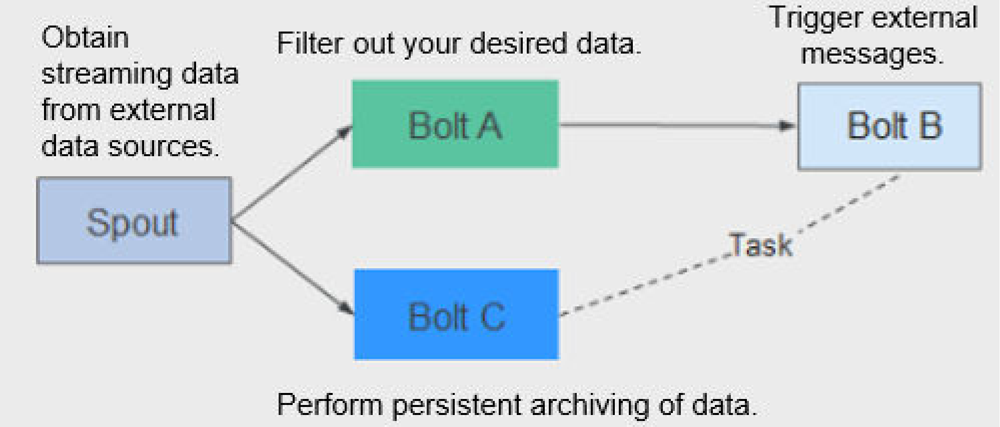
Apache Storm is a distributed, reliable, and fault-tolerant real-time stream data processing system. In Storm, a graph-shaped data structure called topology needs to be designed first for real-time computing. The topology will be submitted to a cluster. Then a master node in the cluster distributes codes and assigns tasks to worker nodes. A topology contains two roles: spout and bolt. A spout sends messages and sends data streams in tuples. A bolt converts the data streams and performs computing and filtering operations. The bolt can randomly send data to other bolts. Tuples sent by a spout are unchangeable arrays and map to fixed key-value pairs.

Figure 1-122 System architecture of Storm



Service processing logic is encapsulated in the topology of Storm. A topology is a set of spout (data sources) and bolt (logical processing) components that are connected using Stream Groupings in DAG mode. All components (spout and bolt) in a topology are working in parallel. In a topology, you can specify the parallelism for each node. Then, Storm allocates tasks in the cluster for computing to improve system processing capabilities.

Figure 1-123 Topology


Obtain streaming data from external data sources.
Filter out your desired data.
Trigger external messages.
Spout
Bolt A
Bolt B
Task
Bolt C
Perform persistent archiving of data.

Storm is applicable to real-time analysis, continuous computing, and distributed extract, transform, and load (ETL). It has the following advantages:

- Wide applications
- High scalability
- Zero data loss
- High fault tolerance
- Easy to construct and control
- Multi-language support

Storm is a computing platform and provides Continuous Query Language (CQL) in the service layer to facilitate service implementation. CQL has the following features:

- Easy to use: The CQL syntax is similar to the SQL syntax. Users who have basic knowledge of SQL can easily learn CQL and use it to develop services.
- Rich functions: In addition to basic expressions provided by SQL, CQL provides functions, such as windows, filtering, and concurrency setting, for stream processing.
- Easy to scale: CQL provides an extension API to support increasingly complex service scenarios. Users can customize the input, output, serialization, and deserialization to meet specific service requirements.
- Easy to debug: CQL provides detailed explanation of error codes, facilitating users to rectify faults.

For details about Storm architecture and principles, see **https://storm.apache.org/**.

Principle

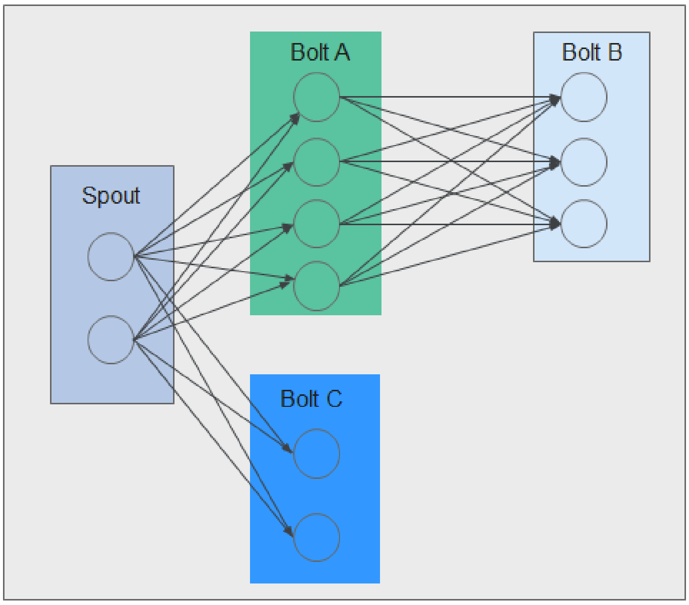
- **Basic Concepts**

Table 1-23 Concepts

Concept	Description
Tuple	A tuple is an invariable key-value pair used to transfer data. Tuples are created and processed in distributed manner.
Stream	A stream is an unbounded sequence of tuples.
Topology	A topology is a real-time application running on the Storm platform. It is a Directed Acyclic Graph (DAG) composed of components. A topology can concurrently run on multiple machines. Each machine runs a part of the DAG. A topology is similar to a MapReduce job. The difference is that the topology is a resident program. Once started, the topology cannot stop unless it is manually terminated.
Spout	A spout is the source of tuples. For example, a spout may read data from a message queue, database, file system, or TCP connection and converts them as tuples, which are processed by the next component.
Bolt	In a Topology, a bolt is a component that receives data and executes specific logic, such as filtering or converting tuples, joining or aggregating streams, and performing statistics and result persistence.
Worker	A Worker is a physical processing in running state in a Topology. Each Worker is a JVM process. Each Topology may be executed by multiple Workers. Each Worker executes a logic subset of the Topology.
Task	A task is a spout or bolt thread of a Worker.
Stream groupings	A stream grouping specifies the tuple dispatching policies. It instructs the subsequent bolt how to receive tuples. The supported policies include Shuffle Grouping, Fields Grouping, All Grouping, Global Grouping, Non Grouping, and Directed Grouping.

Figure 1-124 shows a Topology (DAG) consisting of a Spout and Bolt. In the figure, a rectangle indicates a Spout or Bolt, the node in each rectangle indicate tasks, and the lines between tasks indicate streams.

Figure 1-124 Topology



- **Reliability**

  Storm provides three levels of data reliability:

  - At Most Once: The processed data may be lost, but it cannot be processed repeatedly. This reliability level offers the highest throughput.
  - At Least Once: Data may be processed repeatedly to ensure reliable data transmission. If a response is not received within the specified time, the Spout resends the data to Bolts for processing. This reliability level may slightly affect system performance.
  - Exactly Once: Data is successfully transmitted without loss or redundancy processing. This reliability level delivers the poorest performance.

  Select the reliability level based on service requirements. For example, for the services requiring high data reliability, use Exactly Once to ensure that data is processed only once. For the services insensitive to data loss, use other levels to improve system performance.

- **Fault Tolerance**

  Storm is a fault-tolerant system that offers high availability. **Table 1-24** describes the fault tolerance of the Storm components.

Table 1-24 Fault tolerance

Scenario	Description
Nimbus failed	Nimbus is fail-fast and stateless. If the active Nimbus is faulty, the standby Nimbus takes over services immediately, and provide external services.
Supervisor failed	Supervisor is a background daemon of Workers. It is fail-fast and stateless. If a Supervisor is faulty, the Workers running on the node are not affected but cannot receive new tasks. The OMS can detect the fault of the Supervisor and restart the processes.

Scenario	Description
Worker failed	If a Worker is faulty, the Supervisor on the Worker will restart it again. If the restart fails for multiple times, Nimbus reassigns tasks to other nodes.
Node failed	If a node is faulty, all the tasks being processed by the node time out and Nimbus will assign the tasks to another node for processing.

Open Source Features

- Distributed real-time computing

  In a Storm cluster, each machine supports the running of multiple work processes and each work process can create multiple threads. Each thread can execute multiple tasks. A task indicates concurrent data processing.

- High fault tolerance

  During message processing, if a node or a process is faulty, the message processing unit can be redeployed.

- Reliable messages

  Data processing methods including At-Least Once, At-Most Once, and Exactly Once are supported.

- Security mechanism

  Storm provides Kerberos-based authentication and pluggable authorization mechanisms, supports SSL Storm UI and Log Viewer UI, and supports security integration with other big data platform components (such as ZooKeeper and HDFS).

- Flexible topology defining and deployment

  The Flux framework is used to define and deploy service topologies. If the service DAG is changed, users only need to modify YAML domain specific language (DSL), but do not need to recompile or package service code.

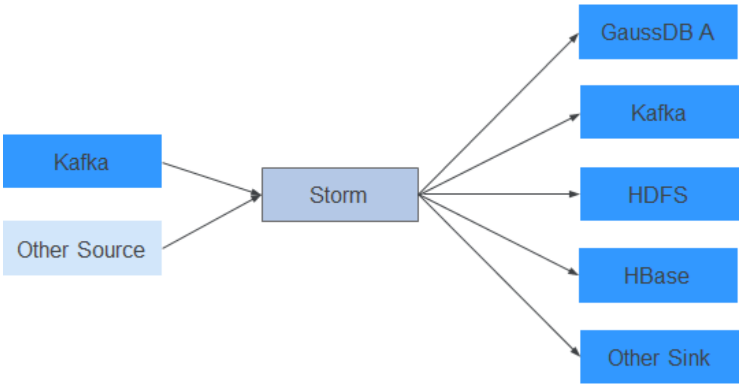
- Integration with external components

  Storm supports integration with multiple external components such as Kafka, HDFS, HBase, Redis, and JDBC/RDBMS, implementing services that involve multiple data sources.

1.3.24.2 Relationship Between Storm and Other Components

Storm provides a real-time distributed computing framework. It can obtain real-time messages from data sources (such as Kafka and TCP connection), perform high-throughput and low-latency real-time computing on a real-time platform, and export results to message queues or implement data persistence. **Figure 1-125** shows the relationship between Storm and other components.

Figure 1-125 Relationship with other components



Relationship between Storm and Streaming

Both Storm and Streaming use the open source Apache Storm kernel. However, the kernel version used by Storm is 1.2.1 whereas that used by Streaming is 0.10.0. Streaming is used to inherit transition services in upgrade scenarios. For example, if Streaming has been deployed in an earlier version and services are running, Streaming can still be used after the upgrade. Storm is recommended in a new cluster.

Storm 1.2.1 has the following new features:

- **Distributed cache**: Provides external resources (configurations) required for sharing and updating the topology using CLI tools. You do not need to re-package and re-deploy the topology.
- **Native Streaming Window API**: Provides window-based APIs.
- **Resource scheduler**: Added the resource scheduler plug-in. When defining a topology, you can specify the maximum resources available and assign resource quotas to users, thus to manage topology resources of the users.
- **State management**: Provides the Bolt API with the checkpoint mechanism. When an event fails, Storm automatically manages the Bolt status and restore the event.
- **Message sampling and debugging**: On the Storm UI, you can enable or disable topology- or component-level debugging to output stream messages to specified logs based on the sampling ratio.
- **Worker dynamic analysis**: On the Storm UI, you can collect jstack and heap logs of the Worker process and restart the Worker process.
- **Dynamic adjustment of topology logs**: You can dynamically change the running topology logs on the CLI or Storm UI.
- **Improved performance**: Compared with earlier versions, the performance of Storm is greatly improved. Although the topology performance is closely related to the use case scenario and dependency on external services, the performance is three times higher in most scenarios.

1.3.24.3 Storm Enhanced Open Source Features

- CQL

  Continuous Query Language (CQL) is an SQL-like language used for real-time stream processing. Compared with SQL, CQL has introduced the concept of (time-sequencing) window, which allows data to be stored and processed in the memory. The CQL output is the computing results of data streams at specific time. The use of CQL accelerates service development, enables tasks to be easily submitted to the Storm platform for real-time processing, facilitates output of results, and allows tasks to be terminated at the appropriate time.

- High Availability

  Nimbus HA ensures continuous service processing such as adding topologies and management even if one Nimbus is faulty, improving cluster availability.

1.3.25 Tez

Tez is Apache's latest open source computing framework that supports Directed Acyclic Graph (DAG) jobs. It can convert multiple dependent jobs into one job, greatly improving the performance of DAG jobs. If projects like Hive and **Pig** use Tez instead of MapReduce as the backbone of data processing, response time will be significantly reduced. Tez is built on YARN and can run MapReduce jobs without any modification.

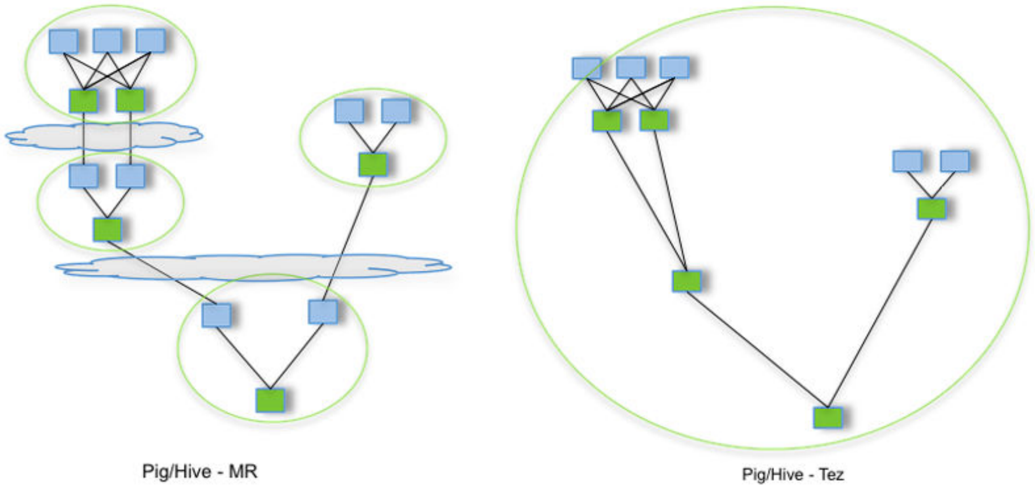
MRS uses Tez as the default execution engine of Hive. Tez remarkably surpasses the original MapReduce computing engine in terms of execution efficiency.

For details about Tez, see **https://tez.apache.org/**.

Relationship Between Tez and MapReduce

Tez uses a DAG to organize MapReduce tasks. In the DAG, a node is an RDD, and an edge indicates an operation on the RDD. The core idea is to further split Map tasks and Reduce tasks. A Map task is split into the Input-Processor-Sort-Merge-Output tasks, and the Reduce task is split into the Input-Shuffle-Sort-Merge-Process-output tasks. Tez flexibly regroups several small tasks to form a large DAG job.

Figure 1-126 Processes for submitting tasks using Hive on MapReduce and Hive on Tez



A Hive on MapReduce task contains multiple MapReduce tasks. Each task stores intermediate results to HDFS. The reducer in the previous step provides data for the mapper in the next step. A Hive on Tez task can complete the same processing process in only one task, and HDFS does not need to be accessed between tasks.

Relationship Between Tez and Yarn

Tez is a computing framework running on Yarn. The runtime environment consists of ResourceManager and ApplicationMaster of Yarn. ResourceManager is a brand new resource manager system, and ApplicationMaster is responsible for cutting MapReduce job data, assigning tasks, applying for resources, scheduling tasks, and tolerating faults. In addition, TezUI depends on TimelineServer provided by Yarn to display the running process of Tez tasks.

1.3.26 Yarn

1.3.26.1 Yarn Basic Principles

The Apache open source community introduces the unified resource management framework **Yarn** to share Hadoop clusters, improve their scalability and reliability, and eliminate a performance bottleneck of JobTracker in the early MapReduce framework.

The fundamental idea of Yarn is to split up the two major functionalities of the JobTracker, resource management and job scheduling/monitoring, into separate daemons. The idea is to have a global ResourceManager (RM) and per-application ApplicationMaster (AM).

NOTE

An application is either a single job in the classical sense of MapReduce jobs or a Directed Acyclic Graph (DAG) of jobs.

Architecture

ResourceManager is the essence of the layered structure of Yarn. This entity controls an entire cluster and manages the allocation of applications to underlying compute resources. The ResourceManager carefully allocates various resources (compute, memory, bandwidth, and so on) to underlying NodeManagers (Yarn's per-node agents). The ResourceManager also works with ApplicationMasters to allocate resources, and works with the NodeManagers to start and monitor their underlying applications. In this context, the ApplicationMaster has taken some of the role of the prior TaskTracker, and the ResourceManager has taken the role of the JobTracker.

ApplicationMaster manages each instance of an application running in Yarn. The ApplicationMaster negotiates resources from the ResourceManager and works with the NodeManagers to monitor container execution and resource usage (CPU and memory resource allocation).

The NodeManager manages each node in a Yarn cluster. The NodeManager provides per-node services in a cluster, from overseeing the management of a container over its lifecycle to monitoring resources and tracking the health of its nodes. MRv1 manages execution of the Map and Reduce tasks through slots, whereas the NodeManager manages abstract containers, which represent per-node resources available for a particular application.

Figure 1-127 Architecture

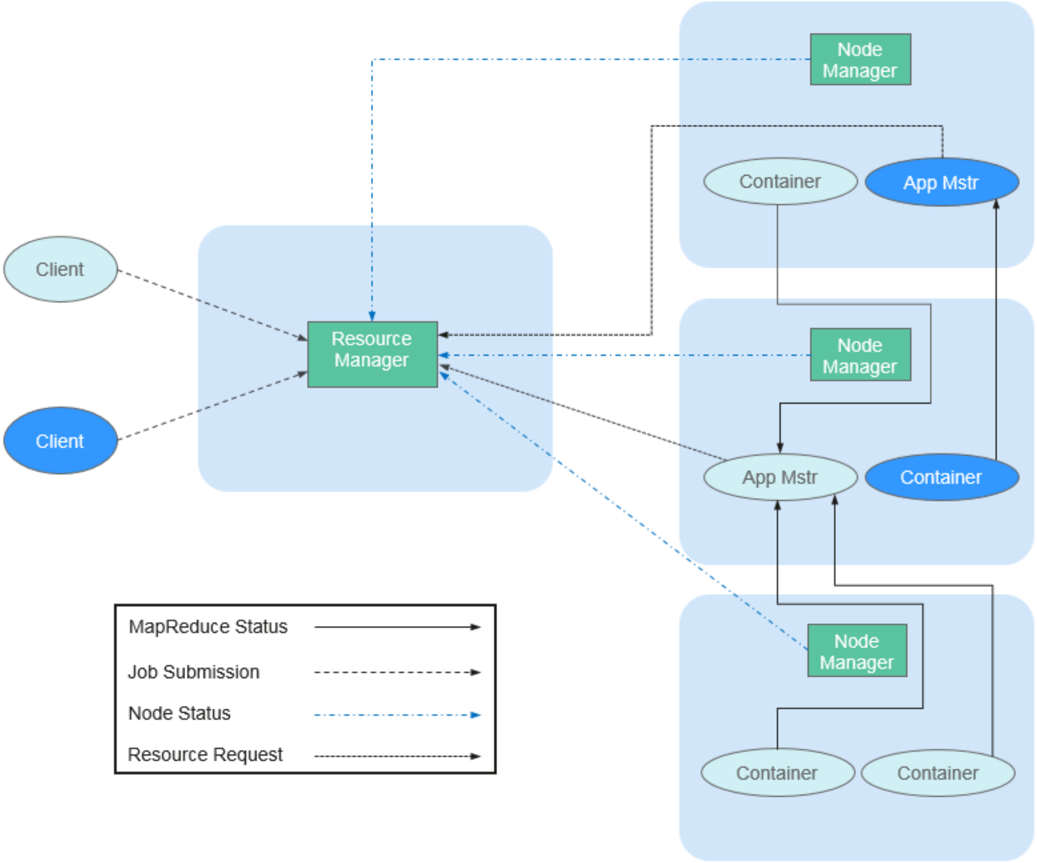


Table 1-25 describes the components shown in **Figure 1-127**.

Table 1-25 Architecture description

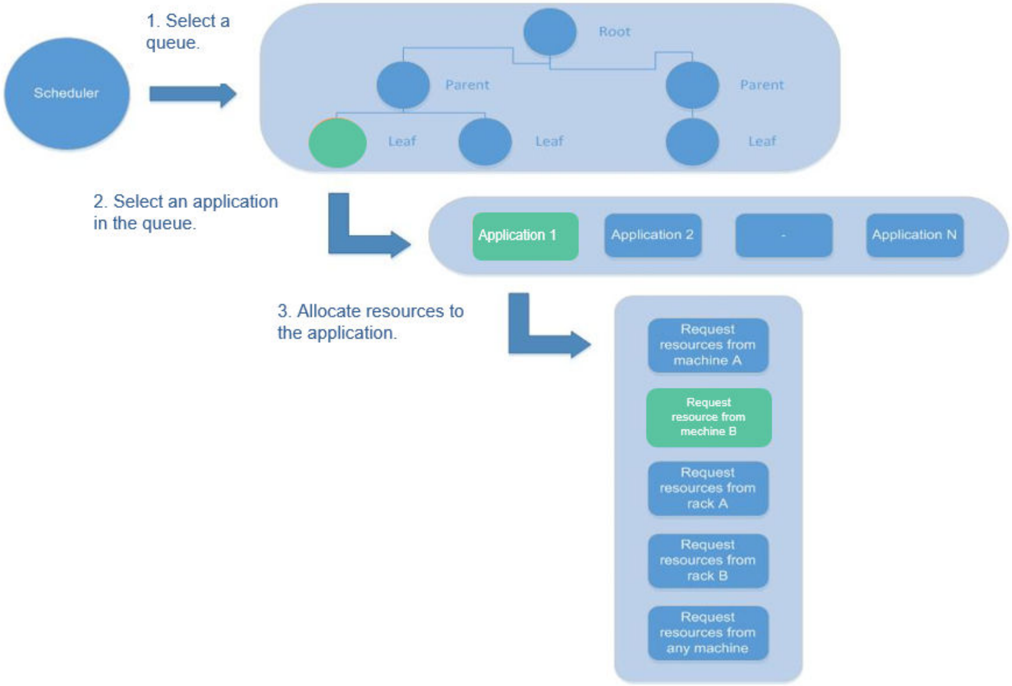
Name	Description
Client	Client of a Yarn application. You can submit a task to ResourceManager and query the operating status of an application using the client.
ResourceManager(RM)	RM centrally manages and allocates all resources in the cluster. It receives resource reporting information from each node (NodeManager) and allocates resources to applications on the basis of the collected resources according a specified policy.
NodeManager(NM)	NM is the agent on each node of Yarn. It manages the computing node in Hadoop cluster, establishes communication with ResourceManger, monitors the lifecycle of containers, monitors the usage of resources such as memory and CPU of each container, traces node health status, and manages logs and auxiliary services used by different applications.

Name	Description
ApplicationMaster(AM)	AM (App Mstr in the figure above) is responsible for all tasks through the lifcycle of in an application. The tasks include the following: Negotiate with an RM scheduler to obtain a resource; further allocate the obtained resources to internal tasks (secondary allocation of resources); communicate with the NM to start or stop tasks; monitor the running status of all tasks; and apply for resources for tasks again to restart the tasks when the tasks fail to be executed.
Container	A resource abstraction in Yarn. It encapsulates multi-dimensional resources (including only memory and CPU) on a certain node. When ApplicationMaster applies for resources from ResourceManager, the ResourceManager returns resources to the ApplicationMaster in a container. Yarn allocates one container for each task and the task can only use the resources encapsulated in the container.

In Yarn, resource schedulers organize resources through hierarchical queues. This ensures that resources are allocated and shared among queues, thereby improving the usage of cluster resources. The core resource allocation model of Superior Scheduler is the same as that of Capacity Scheduler, as shown in the following figure.

A scheduler maintains queue information. You can submit applications to one or more queues. During each NM heartbeat, the scheduler selects a queue according to a specific scheduling rule, selects an application in the queue, and then allocates resources to the application. If resources fail to be allocated to the application due to the limit of some parameters, the scheduler will select another application. After the selection, the scheduler processes the resource request of this application. The scheduler gives priority to the requests for local resources first, and then for resources on the same rack, and finally for resources from any machine.

Figure 1-128 Resource allocation model


1. Select a queue.
Scheduler
Root
Parent
Parent
Leaf
Leaf
Leaf
2. Select an application in the queue.
Application 1
Application 2
-
Application N
3. Allocate resources to the application.
Request resources from machine A
Request resource from mechine B
Request resources from rack A
Request resources from rack B
Request resources from any machine

Principle

The new Hadoop MapReduce framework is named MRv2 or Yarn. Yarn consists of ResourceManager, ApplicationMaster, and NodeManager.

- ResourceManager is a global resource manager that manages and allocates resources in the system. ResourceManager consists of Scheduler and Applications Manager.
  - Scheduler allocates system resources to all running applications based on the restrictions such as capacity and queue (for example, allocates a certain amount of resources for a queue and executes a specific number of jobs). It allocates resources based on the demand of applications, with container being used as the resource allocation unit. Functioning as a dynamic resource allocation unit, Container encapsulates memory, CPU, disk, and network resources, thereby limiting the resource consumed by each task. In addition, the Scheduler is a pluggable component. You can design new schedulers as required. Yarn provides multiple directly available schedulers, such as Fair Scheduler and Capacity Scheduler.
  - Applications Manager manages all applications in the system and involves submitting applications, negotiating with schedulers about resources, enabling and monitoring ApplicationMaster, and restarting ApplicationMaster upon the startup failure.
- NodeManager is the resource and task manager of each node. On one hand, NodeManager periodically reports resource usage of the local node and the running status of each Container to ResourceManager. On the other hand, NodeManager receives and processes requests from ApplicationMaster for starting or stopping Containers.
- ApplicationMaster is responsible for all tasks through the lifecycle of an application, these channels include the following:

- Negotiate with the RM scheduler to obtain resources.
- Assign resources to internal components (secondary allocation of resources).
- Communicates with NodeManager to start or stop tasks.
- Monitor the running status of all tasks, and applies for resources again for tasks when tasks fail to run to restart the tasks.

Capacity Scheduler Principle

Capacity Scheduler is a multi-user scheduler. It allocates resources by queue and sets the minimum/maximum resources that can be used for each queue. In addition, the upper limit of resource usage is set for each user to prevent resource abuse. Remaining resources of a queue can be temporarily shared with other queues.

Capacity Scheduler supports multiple queues. It configures a certain amount of resources for each queue and adopts the first-in-first-out queuing (FIFO) scheduling policy. To prevent one user's applications from exclusively using the resources in a queue, Capacity Scheduler sets a limit on the number of resources used by jobs submitted by one user. During scheduling, Capacity Scheduler first calculates the number of resources required for each queue, and selects the queue that requires the least resources. Then, it allocates resources based on the job priority and time that jobs are submitted as well as the limit on resources and memory. Capacity Scheduler supports the following features:

- Guaranteed capacity: As the administrator, you can set the lower and upper limits of resource usage for each queue. All applications submitted to this queue share the resources.
- High flexibility: Temporarily, the remaining resources of a queue can be shared with other queues. However, such resources must be released in case of new application submission to the queue. Such flexible resource allocation helps notably improve resource usage.
- Multi-tenancy: Multiple users can share a cluster, and multiple applications can run concurrently. To avoid exclusive resource usage by a single application, user, or queue, the administrator can add multiple constraints (for example, limit on concurrent tasks of a single application).
- Assured protection: An ACL list is provided for each queue to strictly limit user access. You can specify the users who can view your application status or control the applications. Additionally, the administrator can specify a queue administrator and a cluster system administrator.
- Dynamic update of configuration files: Administrators can dynamically modify configuration parameters to manage clusters online.

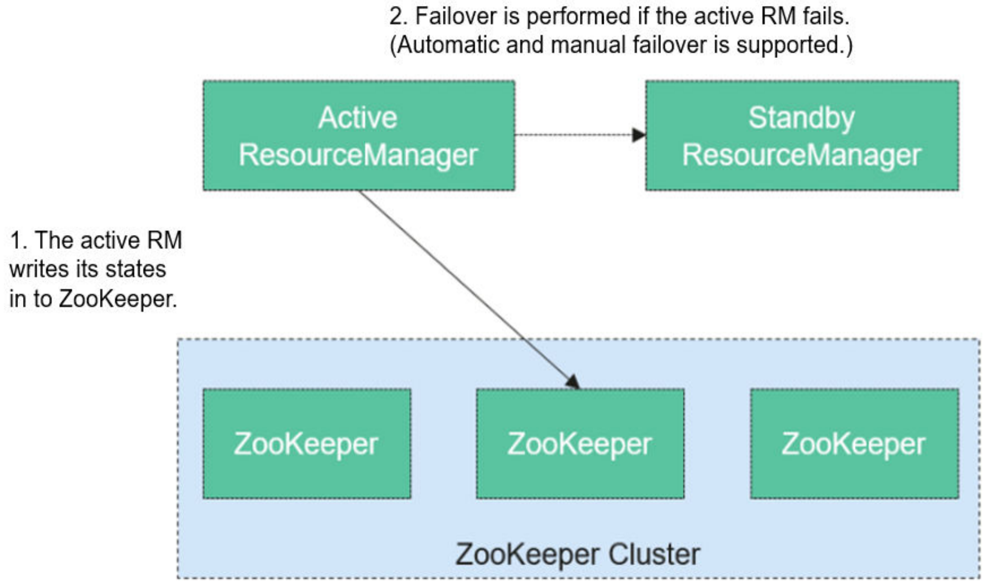
Each queue in Capacity Scheduler can limit the resource usage. However, the resource usage of a queue determines its priority when resources are allocated to queues, indicating that queues with smaller capacity are competitive. If the throughput of a cluster is big, delay scheduling enables an application to give up cross-machine or cross-rack scheduling, and to request local scheduling.

1.3.26.2 Yarn HA Solution

HA Principles and Implementation Solution

ResourceManager in Yarn manages resources and schedules tasks in the cluster. In versions earlier than Hadoop 2.4, SPOFs may occur on ResourceManager in the Yarn cluster. The Yarn HA solution uses redundant ResourceManager nodes to tackle challenges of service reliability and fault tolerance.

Figure 1-129 ResourceManager HA architecture


2. Failover is performed if the active RM fails.
(Automatic and manual failover is supported.)
Active ResourceManager
Standby ResourceManager
1. The active RM writes its states in to ZooKeeper.
ZooKeeper
ZooKeeper
ZooKeeper
ZooKeeper Cluster

ResourceManager HA is achieved using active-standby ResourceManager nodes, as shown in **Figure 1-129**. Similar to the HDFS HA solution, the ResourceManager HA allows only one ResourceManager node to be in the active state at any time. When the active ResourceManager fails, the active-standby switchover can be triggered automatically or manually.

When the automatic failover function is not enabled, after the Yarn cluster is enabled, administrators need to run the ***yarn rmadmin*** command to manually switch one of the ResourceManager nodes to the active state. Upon a planned maintenance event or a fault, they are expected to first demote the active ResourceManager to the standby state and the standby ResourceManager promote to the active state.

When the automatic switchover is enabled, a built-in ActiveStandbyElector that is based on ZooKeeper decide which ResourceManager node should be the active one. When the active ResourceManager is faulty, another ResourceManager node is automatically selected to be the active one to take over the faulty node.

When ResourceManager nodes in the cluster are deployed in HA mode, the configuration **yarn-site.xml** used by clients needs to list all the ResourceManager nodes. The client (including ApplicationMaster and NodeManager) searches for the active ResourceManager in polling mode. That is, the client needs to provide the fault tolerance mechanism. If the active ResourceManager cannot be connected with, the client continuously searches for a new one in polling mode.

After the standby ResourceManager promotes to be the active one, the upper-layer applications can recover to their status when the fault occurs. (For details, see **ResourceManger Restart**.) When ResourceManager Restart is enabled, the restarted ResourceManager node loads the information of the previous active ResourceManager node, and takes over container status information on all NodeManager nodes to continue service running. In this way, status information can be saved by periodically executing checkpoint operations, avoiding data loss. Ensure that both active and standby ResourceManager nodes can access the status information. Currently, three methods are provided for sharing status information by file system (FileSystemRMStateStore), LevelDB database (LeveldbRMStateStore), and ZooKeeper (ZKRMStateStore). Among them, only ZKRMStateStore supports the Fencing mechanism. By default, Hadoop uses ZKRMStateStore.

For more information about the Yarn HA solution, visit the following website:

http://hadoop.apache.org/docs/r3.1.1/hadoop-yarn/hadoop-yarn-site/ResourceManagerHA.html

1.3.26.3 Relationship Between YARN and Other Components

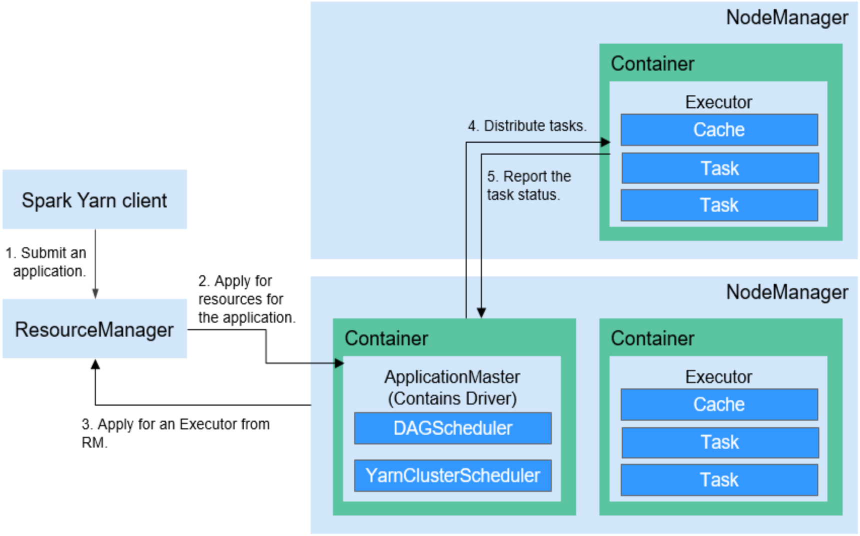
Relationship Between YARN and Spark

The Spark computing and scheduling can be implemented using YARN mode. Spark enjoys the compute resources provided by YARN clusters and runs tasks in a distributed way. Spark on YARN has two modes: YARN-cluster and YARN-client.

- YARN Cluster mode

  Figure 1-130 describes the operation framework.

  Figure 1-130 Spark on YARN-cluster operation framework



Spark on YARN-cluster implementation process:

a. The client generates the application information, and then sends the information to ResourceManager.

b. ResourceManager allocates the first container (ApplicationMaster) to SparkApplication and starts the driver on the container.

c. ApplicationMaster applies for resources from ResourceManager to run the container.

   ResourceManager allocates the containers to ApplicationMaster, which communicates with the related NodeManagers and starts the executor in the obtained container. After the executor is started, it registers with drivers and applies for tasks.

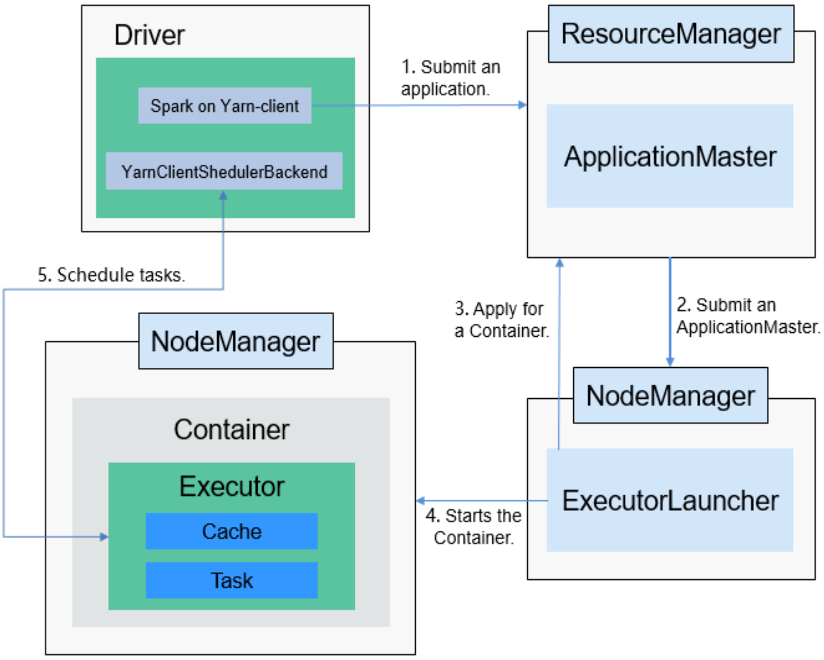
d. Drivers allocate tasks to the executors.

e. Executors run tasks and report the operating status to Drivers.

- YARN Client mode

  Figure 1-131 describes the operation framework.

Figure 1-131 Spark on YARN-client operation framework



Spark on YARN-client implementation process:

NOTE

In YARN-client mode, the driver is deployed and started on the client. In YARN-client mode, the client of an earlier version is incompatible. You are advised to use the YARN-cluster mode.

a. The client sends the Spark application request to ResourceManager, then ResourceManager returns the results. The results include information such as Application ID and the maximum and minimum available resources. The client packages all information required to start ApplicationMaster, and sends the information to ResourceManager.

b. After receiving the request, ResourceManager finds a proper node for ApplicationMaster and starts it on this node. ApplicationMaster is a role in YARN, and the process name in Spark is ExecutorLauncher.

c. Based on the resource requirements of each task, ApplicationMaster can apply for a series of containers to run tasks from ResourceManager.

d. After receiving the newly allocated container list (from ResourceManager), ApplicationMaster sends information to the related NodeManagers to start the containers.

   ResourceManager allocates the containers to ApplicationMaster, which communicates with the related NodeManagers and starts the executor in the obtained container. After the executor is started, it registers with drivers and applies for tasks.

   📖 NOTE

   Running containers are not suspended and resources are not released.

e. Drivers allocate tasks to the executors. Executors run tasks and report the operating status to Drivers.

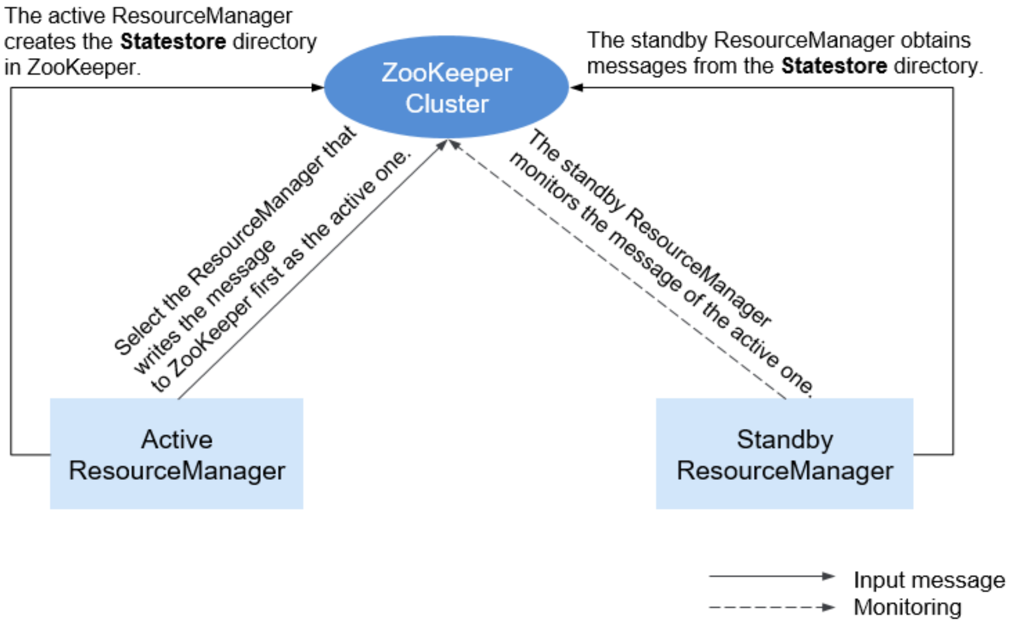
Relationship Between YARN and MapReduce

MapReduce is a computing framework running on YARN, which is used for batch processing. MRv1 is implemented based on MapReduce in Hadoop 1.0, which is composed of programming models (new and old programming APIs), running environment (JobTracker and TaskTracker), and data processing engine (MapTask and ReduceTask). This framework is still weak in scalability, fault tolerance (JobTracker SPOF), and compatibility with multiple frameworks. (Currently, only the MapReduce computing framework is supported.) MRv2 is implemented based on MapReduce in Hadoop 2.0. The source code reuses MRv1 programming models and data processing engine implementation, and the running environment is composed of ResourceManager and ApplicationMaster. ResourceManager is a brand new resource manager system, and ApplicationMaster is responsible for cutting MapReduce job data, assigning tasks, applying for resources, scheduling tasks, and tolerating faults.

Relationship Between YARN and ZooKeeper

Figure 1-132 shows the relationship between ZooKeeper and YARN.

Figure 1-132 Relationship Between ZooKeeper and YARN


The active ResourceManager creates the **Statestore** directory in ZooKeeper.
ZooKeeper Cluster
The standby ResourceManager obtains messages from the **Statestore** directory.
Select the ResourceManager that writes the message to ZooKeeper first as the active one.
The standby ResourceManager monitors the message of the active one.
Active ResourceManager
Standby ResourceManager
Input message
Monitoring

1. When the system is started, ResourceManager attempts to write state information to ZooKeeper. ResourceManager that first writes state information to ZooKeeper is selected as the active ResourceManager, and others are standby ResourceManagers. The standby ResourceManagers periodically monitor active ResourceManager election information in ZooKeeper.
2. The active ResourceManager creates the **Statestore** directory in ZooKeeper to store application information. If the active ResourceManager is faulty, the standby ResourceManager obtains application information from the **Statestore** directory and restores the data.

Relationship Between YARN and Tez

The Hive on Tez job information requires the TimeLine Server capability of YARN so that Hive tasks can display the current and historical status of applications, facilitating storage and retrieval.

1.3.26.4 Yarn Enhanced Open Source Features

Priority-based task scheduling

In the native Yarn resource scheduling mechanism, if the whole Hadoop cluster resources are occupied by those MapReduce jobs submitted earlier, jobs submitted later will be kept in pending state until all running jobs are executed and resources are released.

The MRS cluster provides the task priority scheduling mechanism. With this feature, you can define jobs of different priorities. Jobs of high priority can preempt resources released from jobs of low priority though the high-priority jobs are submitted later. The low-priority jobs that are not started will be suspended unless those jobs of high priority are completed and resources are released, then they can properly be started.

This feature enables services to control computing jobs more flexibly, thereby achieving higher cluster resource utilization.

NOTE

Container reuse is in conflict with task priority scheduling. If container reuse is enabled, resources are being occupied, and task priority scheduling does not take effect.

Yarn Permission Control

The permission mechanism of Hadoop Yarn is implemented through ACLs. The following describes how to grant different permission control to different users:

- Admin ACL

  An O&M administrator is specified for the Yarn cluster. The Admin ACL is determined by **yarn.admin.acl**. The cluster O&M administrator can access the ResourceManager web UI and operate NodeManager nodes, queues, and NodeLabel, **but cannot submit tasks**.

- Queue ACL

  To facilitate user management in the cluster, users or user groups are divided into several queues to which each user and user group belongs. Each queue contains permissions to submit and manage applications (for example, terminate any application).

Open source functions:

Currently, Yarn supports the following roles for users:

- Cluster O&M administrator
- Queue administrator
- Common user

However, the APIs (such as the web UI, REST API, and Java API) provided by Yarn do not support role-specific permission control. Therefore, all users have the permission to access the application and cluster information, which does not meet the isolation requirements in the multi-tenant scenario.

This is an enhanced function.

In security mode, permission management is enhanced for the APIs such as web UI, REST API, and Java API provided by Yarn. Permission control can be performed based on user roles.

Role-based permissions are as follows:

- Cluster O&M administrator: performs management operations in the Yarn cluster, such as accessing the ResourceManager web UI, refreshing queues, setting NodeLabel, and performing active/standby switchover.
- Queue administrator: has the permission to modify and view queues managed by the Yarn cluster.
- Common user: has the permission to modify and view self-submitted applications in the Yarn cluster.

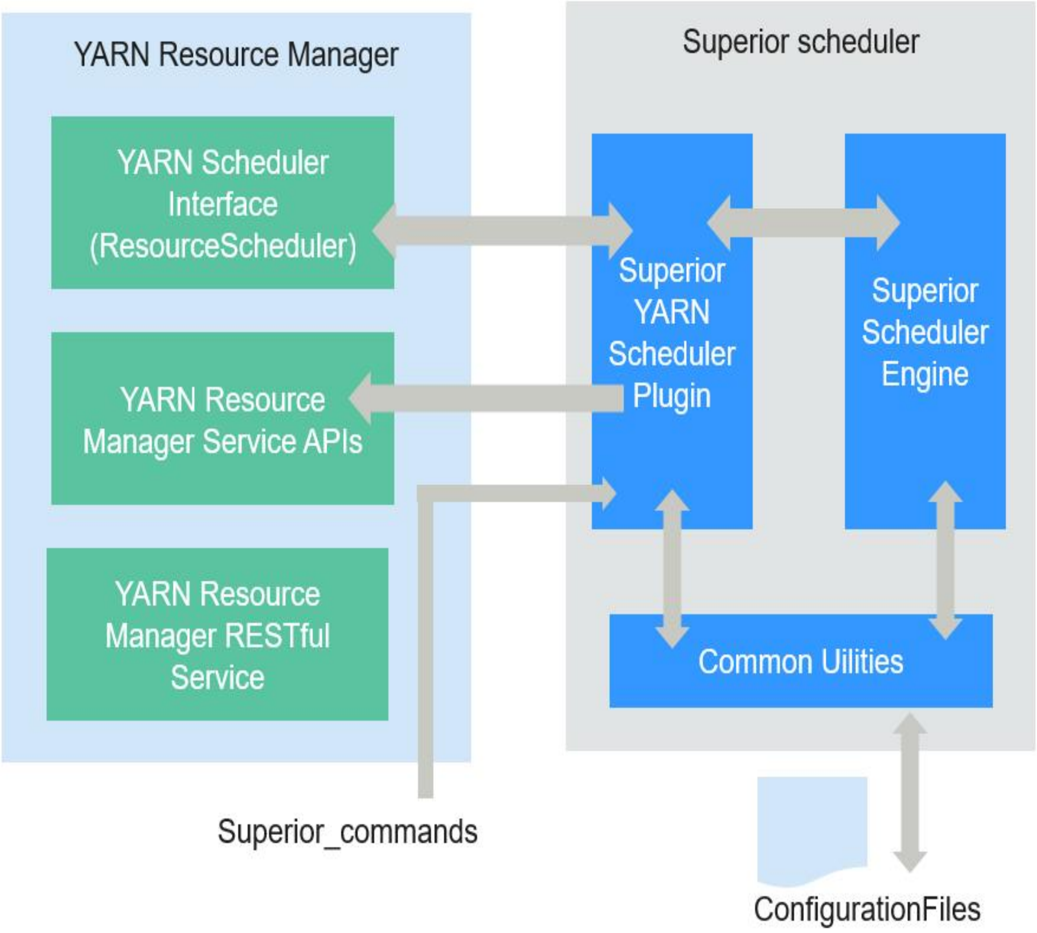
Superior Scheduler Principle (Self-developed)

Superior Scheduler is a scheduling engine designed for the Hadoop Yarn distributed resource management system. It is a high-performance and enterprise-level scheduler designed for converged resource pools and multi-tenant service requirements.

Superior Scheduler achieves all functions of open source schedulers, Fair Scheduler, and Capacity Scheduler. Compared with the open source schedulers, Superior Scheduler is enhanced in the enterprise multi-tenant resource scheduling policy, resource isolation and sharing among users in a tenant, scheduling performance, system resource usage, and cluster scalability. Superior Scheduler is designed to replace open source schedulers.

Similar to open source Fair Scheduler and Capacity Scheduler, Superior Scheduler follows the Yarn scheduler plugin API to interact with Yarn ResourceManager to offer resource scheduling functionalities. **Figure 1-133** shows the overall system diagram.

Figure 1-133 Internal architecture of Superior Scheduler


YARN Resource Manager
YARN Scheduler Interface (ResourceScheduler)
YARN Resource Manager Service APIs
YARN Resource Manager RESTful Service
Superior scheduler
Superior YARN Scheduler Plugin
Superior Scheduler Engine
Common Uilities
Superior_commands
ConfigurationFiles

In **Figure 1-133**, Superior Scheduler consists of the following modules:

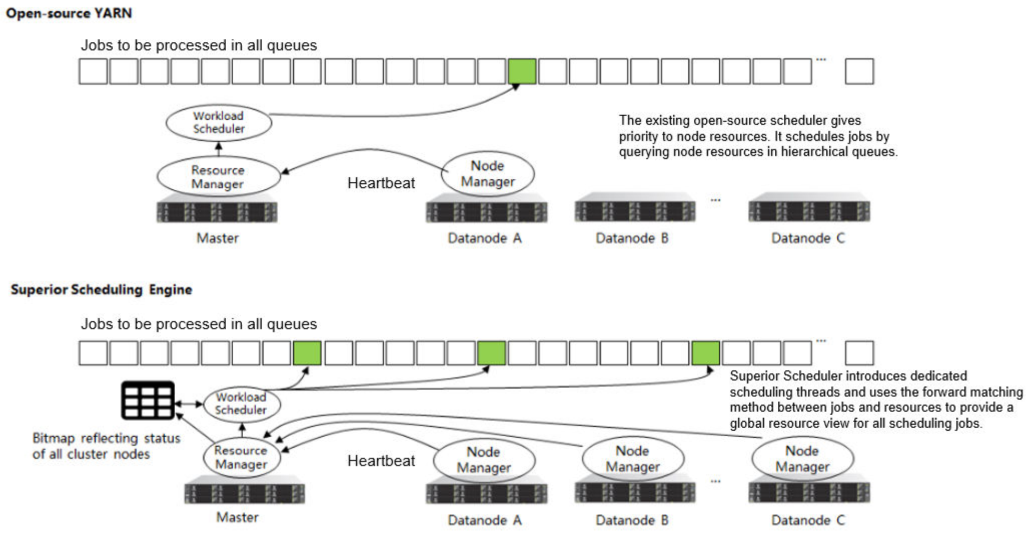
- Superior Scheduler Engine is a high performance scheduler engine with rich scheduling policies.

- Superior Yarn Scheduler Plugin functions as a bridge between Yarn ResourceManager and Superior Scheduler Engine and interacts with Yarn ResourceManager.

  The scheduling principle of open source schedulers is that resources match jobs based on the heartbeats of computing nodes. Specifically, each computing node periodically sends heartbeat messages to ResourceManager of Yarn to notify the node status and starts the scheduler to assign jobs to the node itself. In this scheduling mechanism, the scheduling period depends on the heartbeat. If the cluster scale increases, bottleneck on system scalability and scheduling performance may occur. In addition, because resources match jobs, the scheduling accuracy of an open source scheduler is limited. For example, data affinity is random and the system does not support load-based scheduling policies. The scheduler may not make the best choice due to lack of the global resource view when selecting jobs.

  Superior Scheduler adopts multiple scheduling mechanisms. There are dedicated scheduling threads in Superior Scheduler, separating heartbeats with scheduling and preventing system heartbeat storms. Additionally, Superior Scheduler matches jobs with resources, providing each scheduled job with a global resource view and increasing the scheduling accuracy. Compared with the open source scheduler, Superior Scheduler excels in system throughput, resource usage, and data affinity.

Figure 1-134 Comparison of Superior Scheduler with open source schedulers



The existing open-source scheduler gives priority to node resources. It schedules jobs by querying node resources in hierarchical queues.

Superior Scheduler introduces dedicated scheduling threads and uses the forward matching method between jobs and resources to provide a global resource view for all scheduling jobs.

Apart from the enhanced system throughput and utilization, Superior Scheduler provides following major scheduling features:

- Multiple resource pools

  Multiple resource pools help logically divide cluster resources and share them among multiple tenants or queues. The division of resource pools supports heterogeneous resources. Resource pools can be divided exactly according to requirements on the application resource isolation. You can configure further policies for different queues in a pool.
- Multi-tenant scheduling (**reserve**, **min**, **share**, and **max**) in each resource pool

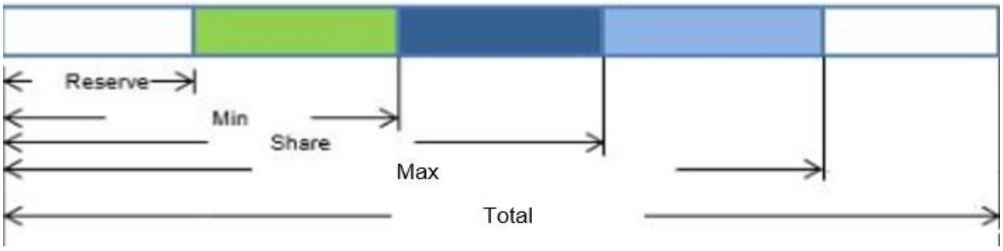
Superior Scheduler provides flexible hierarchical multi-tenant scheduling policy. Different policies can be configured for different tenants or queues that can access different resource pools. The following figure lists supported policies:

Table 1-26 Policy description

Name	Description
reserve	This policy is used to reserve resources for a tenant. Even though tenant has no jobs available, other tenant cannot use the reserved resource. The value can be a percentage or an absolute value. If both the percentage and absolute value are configured, the percentage is automatically calculated into an absolute value, and the larger value is used. The default **reserve** value is **0**. Compared with the method of specifying a dedicated resource pool and hosts, the **reserve** policy provides a flexible floating reservation function. In addition, because no specific hosts are specified, the data affinity for calculation is improved and the impact by the faulty hosts is avoided.
min	This policy allows preemption of minimum resources. Other tenants can use these resources, but the current tenant has the priority to use them. The value can be a percentage or an absolute value. If both the percentage and absolute value are configured, the percentage is automatically calculated into an absolute value, and the larger value is used. The default value is **0**.
share	This policy is used for shared resources that cannot be preempted. To use these resources, the current tenant needs to wait for other tenants to complete jobs and release resources. The value can be a percentage or an absolute value.
max	This policy is used for the maximum resources that can be utilized. The tenant cannot obtain more resources than the allowed maximum value. The value can be a percentage or an absolute value. If both the percentage and absolute value are configured, the percentage is automatically calculated into an absolute value, and the larger value is used. By default value, there is no restriction on resources.

Figure 1-135 shows the tenant resource allocation policy.

Figure 1-135 Resource scheduling policies



NOTE

In the above figure, **Total** indicates the total number of resources, not the scheduling policy.

Compared with open source schedulers, Superior Scheduler supports both percentage and absolute value of tenants for allocating resources, flexibly addressing resource scheduling requirements of enterprise-level tenants. For example, resources can be allocated according to the absolute value of level-1 tenants, avoiding impact caused by changes of cluster scale. However, resources can be allocated according to the allocation percentage of sub-tenants, improving resource usages in the level-1 tenant.

- Heterogeneous and multi-dimensional resource scheduling

  Superior Scheduler supports following functions except CPU and memory scheduling:

  - **Node labels** can be used to identify multi-dimensional attributes of nodes such as **GPU_ENABLED** and **SSD_ENBALED**, and can be scheduled based on these labels.
  - Resource pools can be used to group resources of the same type and allocate them to specific tenants or queues.

- Fair scheduling of multiple users in a tenant

  In a leaf tenant, multiple users can use the same queue to submit jobs. Compared with the open source schedulers, Superior Scheduler supports configuring flexible resource sharing policy among different users in a same tenant. For example, VIP users can be configured with higher resource access weight.

- Data locality aware scheduling

  Superior Scheduler adopts the job-to-node scheduling policy. That is, Superior Scheduler attempts to schedule specified jobs between available nodes so that the selected node is suitable for the specified jobs. By doing so, the scheduler will have an overall view of the cluster and data. Localization is ensured if there is an opportunity to place tasks closer to the data. The open source scheduler uses the node-to-job scheduling policy to match the appropriate jobs to a given node.

- Dynamic resource reservation during container scheduling

  In a heterogeneous and diversified computing environment, some containers need more resources or multiple resources. For example, Spark job may require large memory. When such containers compete with containers requiring fewer resources, containers requiring more resources may not obtain sufficient resources within a reasonable period. Open source schedulers allocate resources to jobs, which may cause unreasonable resource reservation for these jobs. This mechanism leads to the waste of overall system resources. Superior Scheduler differs from open source schedulers in following aspects:

   - Requirement-based matching: Superior Scheduler schedules jobs to nodes and selects appropriate nodes to reserve resources to improve the startup time of containers and avoid waste.
   - Tenant rebalancing: When the reservation logic is enabled, the open source schedulers do not comply with the configured sharing policy. Superior Scheduler uses different methods. In each scheduling period, Superior Scheduler traverses all tenants and attempts to balance resources based on the multi-tenant policy. In addition, Superior Scheduler attempts to meet all policies (**reserve**, **min**, and **share**) to release reserved resources and direct available resources to other containers that should obtain resources under different tenants.

- Dynamic queue status control (**Open**/**Closed**/**Active**/**Inactive**)

  Multiple queue statuses are supported, helping administrators operate and maintain multiple tenants.

   - Open status (**Open/Closed**): If the status is **Open** by default, applications submitted to the queue are accepted. If the status is **Closed**, no application is accepted.
   - Active status (**Active/Inactive**): If the status is **Active** by default, resources can be scheduled and allocated to applications in the tenant. Resources will not be scheduled to queues in **Inactive** status.

- Application pending reason

  If the application is not started, provide the job pending reasons.

Table 1-27 describes the comparison result of Superior Scheduler and Yarn open source schedulers.

Table 1-27 Comparative analysis

Scheduling	Yarn Open Source Scheduler	Superior Scheduler
Multi-tenant scheduling	In homogeneous clusters, either Capacity Scheduler or Fair Scheduler can be selected and the cluster does not support Fair Scheduler. Capacity Scheduler supports the scheduling by percentage and Fair Scheduler supports the scheduling by absolute value.	● Supports heterogeneous clusters and multiple resource pools. ● Supports **reservation** to ensure direct access to resources.
Data locality aware scheduling	The node-to-job scheduling policy reduces the success rate of data localization and potentially affects application execution performance.	The **job-to-node scheduling policy** can aware data location more accurately, and the job hit rate of data localization scheduling is higher.

Scheduling	Yarn Open Source Scheduler	Superior Scheduler
Balanced scheduling based on load of hosts	Not supported	**Balanced scheduling can be achieved when Superior Scheduler considers the host load and resource allocation during scheduling.**
Fair scheduling of multiple users in a tenant	Not supported	Supports keywords **default** and **others**.
Job waiting reason	Not supported	Job waiting reasons illustrate why a job needs to wait.

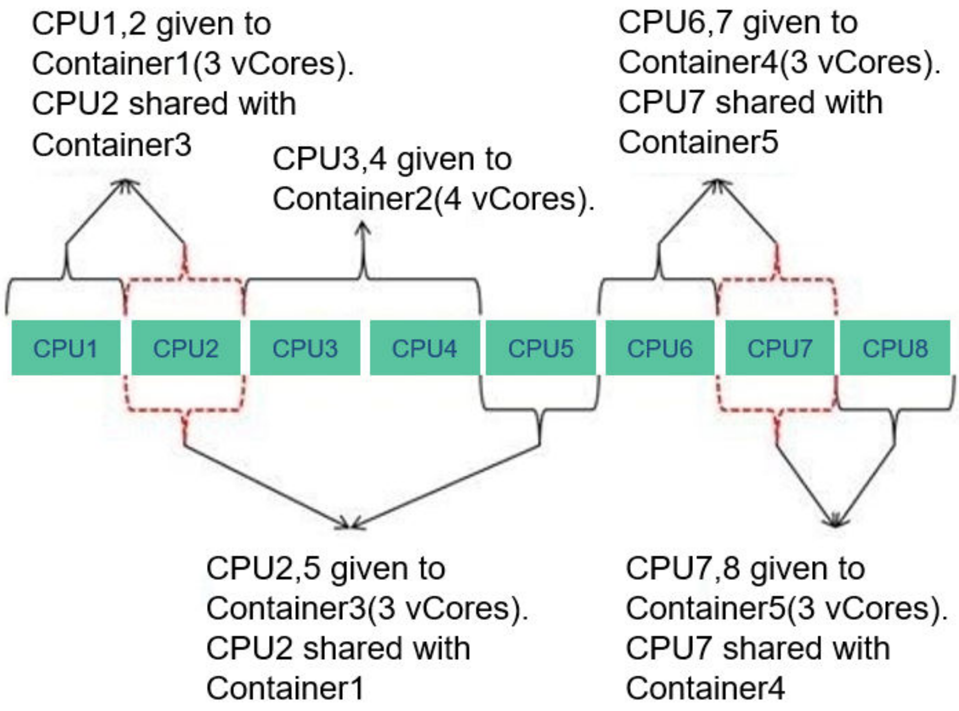
In conclusion, Superior Scheduler is a high-performance scheduler with various scheduling policies and is better than Capacity Scheduler in terms of functionality, performance, resource usage, and scalability.

CPU Hard Isolation

Yarn cannot strictly control the CPU resources used by each container. When the CPU subsystem is used, a container may occupy excessive resources. Therefore, CPUset is used to control resource allocation.

To solve this problem, the CPU resources are allocated to each container based on the ratio of virtual cores (vCores) to physical cores. If a container requires an entire physical core, the container has it. If a container needs only some physical cores, several containers may share the same physical core. The following figure shows an example of the CPU quota. The given ratio of vCores to physical cores is 2:1.

Figure 1-136 CPU quota

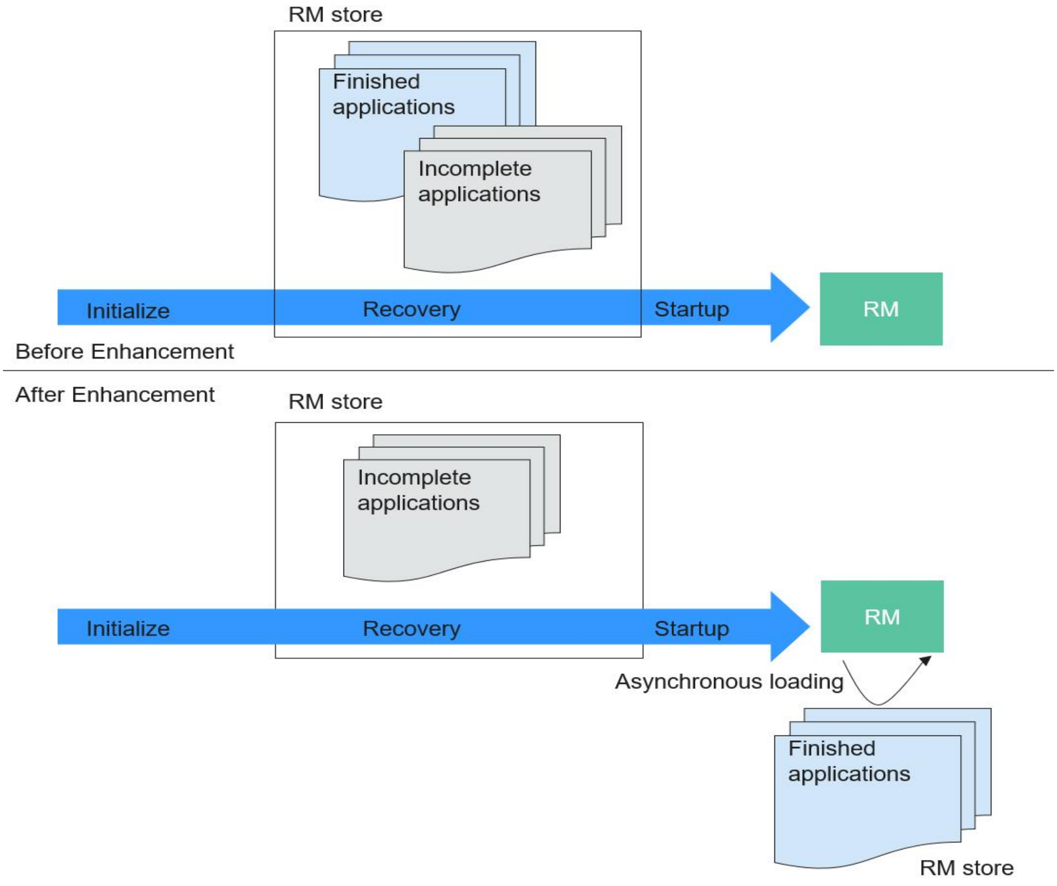

CPU1,2 given to Container1(3 vCores). CPU2 shared with Container3
CPU3,4 given to Container2(4 vCores).
CPU6,7 given to Container4(3 vCores). CPU7 shared with Container5
CPU1
CPU2
CPU3
CPU4
CPU5
CPU6
CPU7
CPU8
CPU2,5 given to Container3(3 vCores). CPU2 shared with Container1
CPU7,8 given to Container5(3 vCores). CPU7 shared with Container4

Enhanced Open Source Feature: Optimizing Restart Performance

Generally, the recovered ResourceManager can obtain running and completed applications. However, a large number of completed applications may cause problems such as slow startup and long HA switchover/restart time of ResourceManagers.

To speed up the startup, obtain the list of unfinished applications before starting the ResourceManagers. In this case, the completed application continues to be recovered in the background asynchronous thread. The following figure shows how the ResourceManager recovery starts.

Figure 1-137 Starting the ResourceManager recovery



1.3.27 ZooKeeper

1.3.27.1 ZooKeeper Basic Principle

ZooKeeper Overview

ZooKeeper is a distributed, highly available coordination service. ZooKeeper provides two functions:

- Prevents the system from single point of failures (SPOFs) and provides reliable services for applications.
- Provides distributed coordination services and manages configuration information.

ZooKeeper Architecture

Nodes in a ZooKeeper cluster have three roles: Leader, Follower, and Observer. **Figure 1-138** shows the ZooKeeper architecture. Generally, an odd number (2N+1) of ZooKeeper servers are configured. At least (N+1) vote majority is required to successfully perform write operation.

Figure 1-138 ZooKeeper architecture

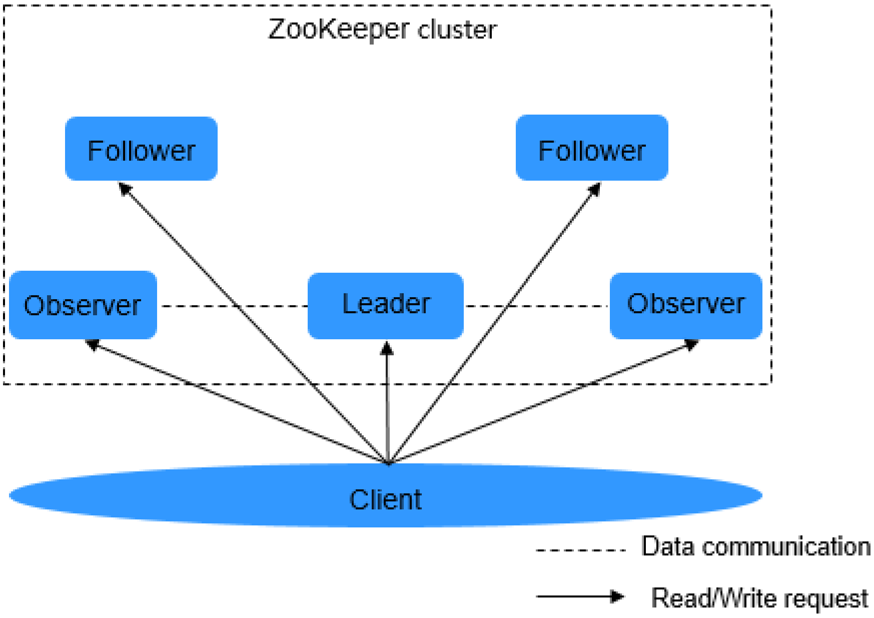


Table 1-28 describes the functions of each module shown in **Figure 1-138**.

Table 1-28 ZooKeeper modules

Module	Description
Leader	Only one node serves as the Leader in a ZooKeeper cluster. The Leader, elected by Followers using the ZooKeeper Atomic Broadcast (ZAB) protocol, receives and coordinates all write requests and synchronizes written information to Followers and Observers.
Follower	Follower has two functions: ● Prevents SPOF. A new Leader is elected from Followers when the Leader is faulty. ● Processes read requests and interacts with the Leader to process write requests.
Observer	The Observer does not take part in voting for election and write requests. It only processes read requests and forwards write requests to the Leader, increasing system processing efficiency.
Client	Reads and writes data from or to the ZooKeeper cluster. For example, HBase can serve as a ZooKeeper client and use the arbitration function of the ZooKeeper cluster to control the active/standby status of the HMaster.

If security services are enabled in the cluster, authentication is required during the connection to ZooKeeper. The authentication modes are as follows:

- keytab mode: Obtain a human-machine user from the administrator for login to the platform and authentication, and obtain the keytab file of the user.
- Ticket mode: Obtain a human-machine user from the administrator for subsequent secure login, enable the renewable and forwardable functions of the Kerberos service, set the ticket update interval, and restart Kerberos and related components.

NOTE

- The default validity period of a user password is 90 days. Therefore, the validity period of the obtained keytab file is 90 days. To prolong the validity period of the keytab file, modify the user password policy and obtain the keytab file again. For details, see the *Administrator Guide*.
- The parameters for enabling the renewable and forwardable functions and setting the ticket update interval are on the **System** tab of the Kerberos service configuration page. The ticket update interval can be set to **kdc_renew_lifetime** or **kdc_max_renewable_life** based on the actual situation.

ZooKeeper Principle

- **Write Request**
  a. After the Follower or Observer receives a write request, the Follower or Observer sends the request to the Leader.
  b. The Leader coordinates Followers to determine whether to accept the write request by voting.
  c. If more than half of voters return a write success message, the Leader submits the write request and returns a success message. Otherwise, a failure message is returned.
  d. The Follower or Observer returns the processing results.
- **Read Request**

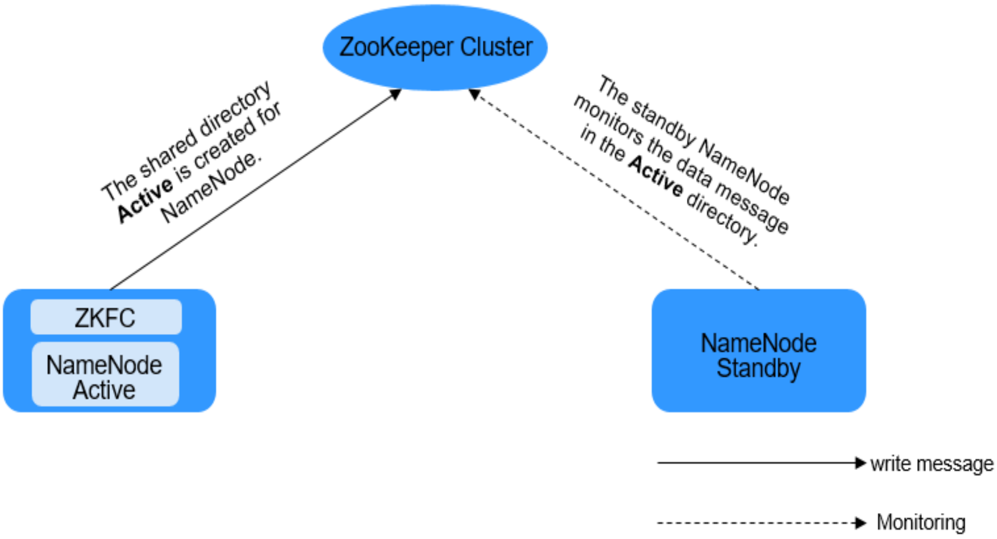
  The client directly reads data from the Leader, Follower, or Observer.

1.3.27.2 Relationship Between ZooKeeper and Other Components

Relationship Between ZooKeeper and HDFS

Figure 1-139 shows the relationship between ZooKeeper and HDFS.

Figure 1-139 Relationship between ZooKeeper and HDFS


ZooKeeper Cluster
The shared directory **Active** is created for NameNode.
The standby NameNode monitors the data message in the **Active** directory.
ZKFC
NameNode Active
NameNode Standby
write message
Monitoring

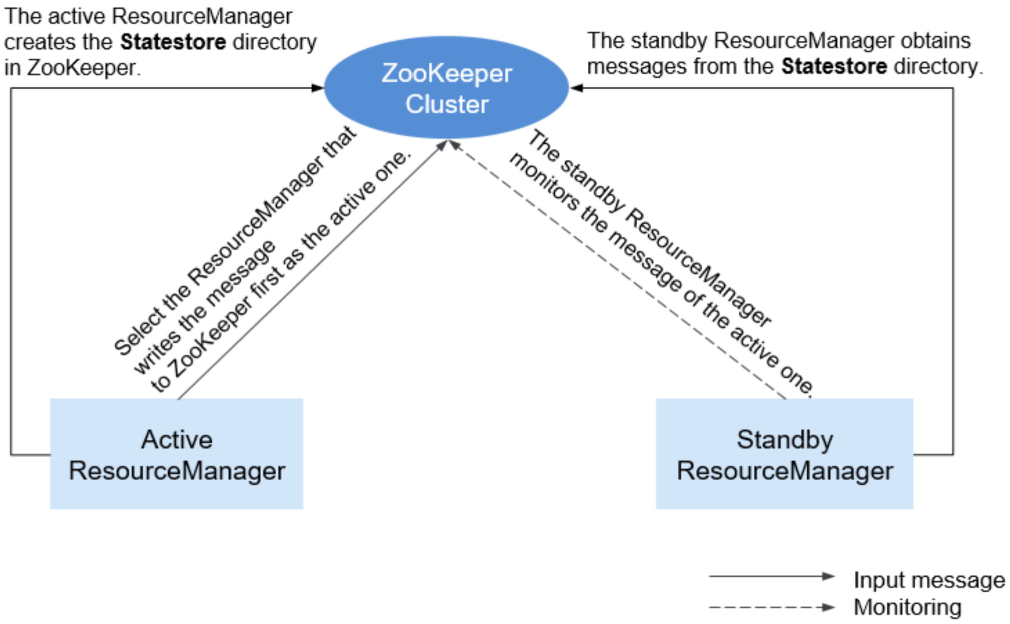
As the client of a ZooKeeper cluster, ZKFailoverController (ZKFC) monitors the status of NameNode. ZKFC is deployed only in the node where NameNode resides, and in both the active and standby HDFS NameNodes.

1. The ZKFC connects to ZooKeeper and saves information such as host names to ZooKeeper under the znode directory **/hadoop-ha**. NameNode that creates the directory first is considered as the active node, and the other is the standby node. NameNodes read the NameNode information periodically through ZooKeeper.
2. When the process of the active node ends abnormally, the standby NameNode detects changes in the **/hadoop-ha** directory through ZooKeeper, and then takes over the service of the active NameNode.

Relationship Between ZooKeeper and YARN

Figure 1-140 shows the relationship between ZooKeeper and YARN.

Figure 1-140 Relationship Between ZooKeeper and YARN

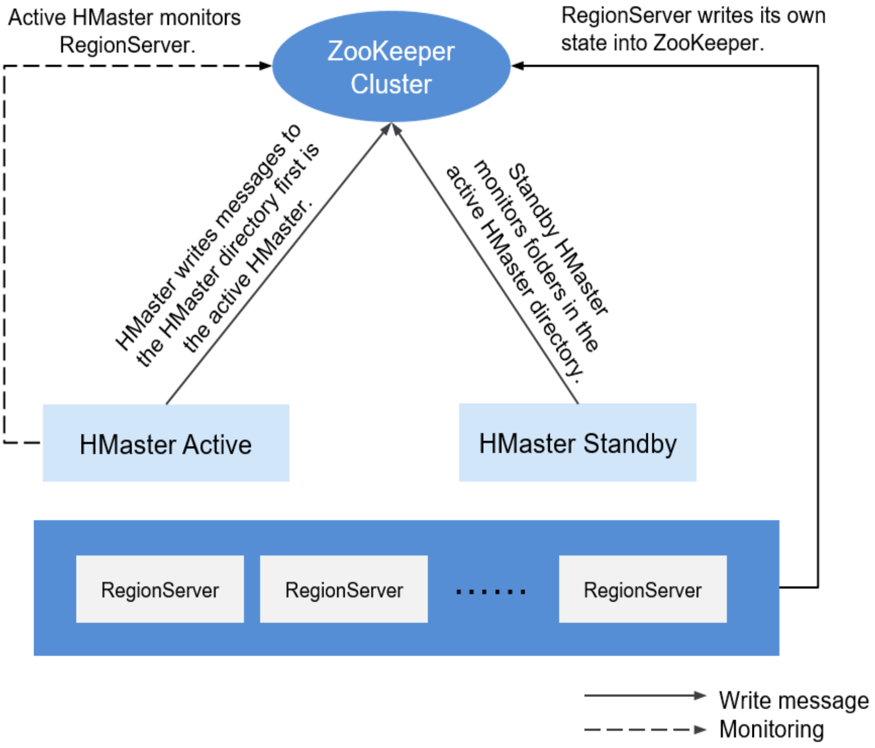

The active ResourceManager creates the **Statestore** directory in ZooKeeper.
ZooKeeper Cluster
The standby ResourceManager obtains messages from the **Statestore** directory.
Select the ResourceManager that writes the message to ZooKeeper first as the active one.
The standby ResourceManager monitors the message of the active one.
Active ResourceManager
Standby ResourceManager
Input message
Monitoring

1. When the system is started, ResourceManager attempts to write state information to ZooKeeper. ResourceManager that first writes state information to ZooKeeper is selected as the active ResourceManager, and others are standby ResourceManagers. The standby ResourceManagers periodically monitor active ResourceManager election information in ZooKeeper.
2. The active ResourceManager creates the **Statestore** directory in ZooKeeper to store application information. If the active ResourceManager is faulty, the standby ResourceManager obtains application information from the **Statestore** directory and restores the data.

Relationship Between ZooKeeper and HBase

Figure 1-141 shows the relationship between ZooKeeper and HBase.

Figure 1-141 Relationship between ZooKeeper and HBase

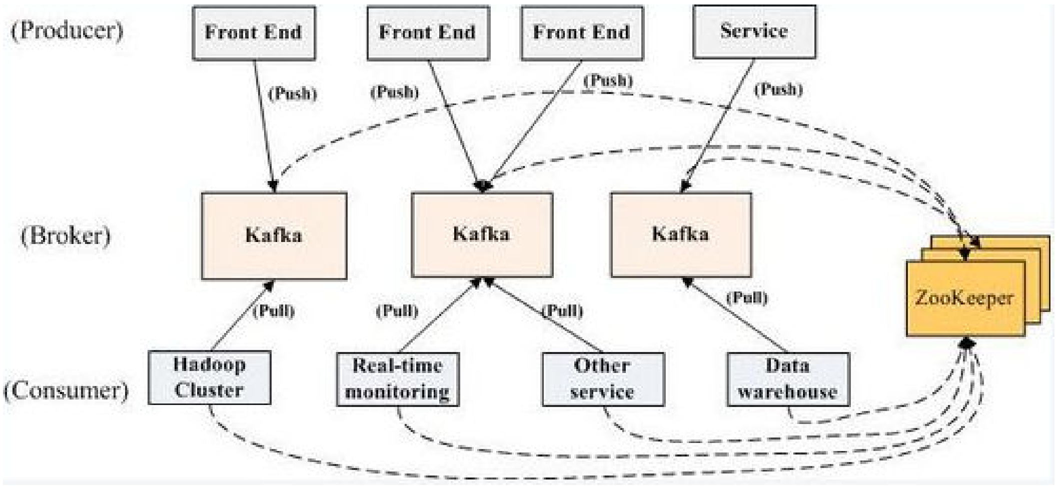

Active HMaster monitors RegionServer.
ZooKeeper Cluster
RegionServer writes its own state into ZooKeeper.
HMaster writes messages to the HMaster directory first is the active HMaster.
Standby HMaster monitors folders in the active HMaster directory.
HMaster Active
HMaster Standby
RegionServer
RegionServer
......
RegionServer
Write message
Monitoring

1. HRegionServer registers itself to ZooKeeper on Ephemeral node. ZooKeeper stores the HBase information, including the HBase metadata and HMaster addresses.
2. HMaster detects the health status of each HRegionServer using ZooKeeper, and monitors them.
3. HBase supports multiple HMaster nodes (like HDFS NameNodes). When the active HMatser is faulty, the standby HMaster obtains the state information about the entire cluster using ZooKeeper. That is, using ZooKeeper can avoid HBase SPOFs.

Relationship Between ZooKeeper and Kafka

Figure 1-142 shows the relationship between ZooKeeper and Kafka.

Figure 1-142 Relationship between ZooKeeper and Kafka



1. Broker uses ZooKeeper to register broker information and elect a partition leader.
2. The consumer uses ZooKeeper to register consumer information, including the partition list of consumer. In addition, ZooKeeper is used to discover the broker list, establish a socket connection with the partition leader, and obtain messages.

1.3.27.3 ZooKeeper Enhanced Open Source Features

Enhanced Log

In security mode, an ephemeral node is deleted as long as the session that created the node expires. Ephemeral node deletion is recorded in audit logs so that ephemeral node status can be obtained.

Usernames must be added to audit logs for all operations performed on ZooKeeper clients.

On the ZooKeeper client, create a znode, of which the Kerberos principal is **zkcli/hadoop.***<System domain name>***@***<System domain name>*.

For example, open the **<ZOO_LOG_DIR>/zookeeper_audit.log** file. The file content is as follows:

```
2016-12-28 14:17:10,505 | INFO  | CommitProcWorkThread-4 | session=0x12000007553b4903?
user=10.177.223.78,zkcli/hadoop.hadoop.com@HADOOP.COM?ip=10.177.223.78?operation=create znode?
target=ZooKeeperServer?znode=/test1?result=success
2016-12-28 14:17:10,530 | INFO  | CommitProcWorkThread-4 | session=0x12000007553b4903?
user=10.177.223.78,zkcli/hadoop.hadoop.com@HADOOP.COM?ip=10.177.223.78?operation=create znode?
target=ZooKeeperServer?znode=/test2?result=success
2016-12-28 14:17:10,550 | INFO  | CommitProcWorkThread-4 | session=0x12000007553b4903?
user=10.177.223.78,zkcli/hadoop.hadoop.com@HADOOP.COM?ip=10.177.223.78?operation=create znode?
target=ZooKeeperServer?znode=/test3?result=success
2016-12-28 14:17:10,570 | INFO  | CommitProcWorkThread-4 | session=0x12000007553b4903?
user=10.177.223.78,zkcli/hadoop.hadoop.com@HADOOP.COM?ip=10.177.223.78?operation=create znode?
target=ZooKeeperServer?znode=/test4?result=success
2016-12-28 14:17:10,592 | INFO  | CommitProcWorkThread-4 | session=0x12000007553b4903?
user=10.177.223.78,zkcli/hadoop.hadoop.com@HADOOP.COM?ip=10.177.223.78?operation=create znode?
target=ZooKeeperServer?znode=/test5?result=success
2016-12-28 14:17:10,613 | INFO  | CommitProcWorkThread-4 | session=0x12000007553b4903?
user=10.177.223.78,zkcli/hadoop.hadoop.com@HADOOP.COM?ip=10.177.223.78?operation=create znode?
```

```
target=ZooKeeperServer?znode=/test6?result=success
2016-12-28 14:17:10,633 | INFO  | CommitProcWorkThread-4 | session=0x12000007553b4903?
user=10.177.223.78,zkcli/hadoop.hadoop.com@HADOOP.COM?ip=10.177.223.78?operation=create znode?
target=ZooKeeperServer?znode=/test7?result=success
```

The content shows that logs of the ZooKeeper client user **zkcli/hadoop.hadoop.com@HADOOP.COM** are added to the audit log.

User details in ZooKeeper

In ZooKeeper, different authentication schemes use different credentials as users. Based on the authentication provider requirement, any parameter can be considered as users.

Example:

- **SAMLAuthenticationProvider** uses the client principal as a user.
- **X509AuthenticationProvider** uses the user client certificate as a user.
- **IAuthenticationProvider** uses the client IP address as a user.
- A username can be obtained from the custom authentication provider by implementing the **org.apache.zookeeper.server.auth.ExtAuthenticationProvider.getUserName(String)** method. If the method is not implemented, getting the username from the authentication provider instance will be skipped.

Enhanced Open Source Feature: ZooKeeper SSL Communication (Netty Connection)

The ZooKeeper design contains the Nio package and does not support SSL later than version 3.5. To solve this problem, Netty is added to ZooKeeper. Therefore, if you need to use SSL, enable Netty and set the following parameters on the server and client:

The open source server supports only plain text passwords, which may cause security problems. Therefore, such text passwords are no longer used on the server.

- Client
  a. Set **-Dzookeeper.client.secure** in the **zkCli.sh/zkEnv.sh** file to **true** to use secure communication on the client. Then, the client can connect to the secureClientPort on the server.
  b. Set the following parameters in the **zkCli.sh/zkEnv.sh** file to configure the client environment:

Parameter	Description
-Dzookeeper.clientCnxnSocket	Used for Netty communication between clients. Default value: **org.apache.zookeeper.ClientCnxnSocketNetty**
-Dzookeeper.ssl.keyStore.location	Indicates the path for storing the keystore file.

Parameter	Description
-Dzookeeper.ssl.keyStore.password	Encrypts a password.
-Dzookeeper.ssl.trustStore.location	Indicates the path for storing the truststore file.
-Dzookeeper.ssl.trustStore.password	Encrypts a password.
-Dzookeeper.config.crypt.class	Decrypts an encrypted password.
-Dzookeeper.ssl.password.encrypted	Default value: **false** If the keystore and truststore passwords are encrypted, set this parameter to **true**.
-Dzookeeper.ssl.enabled.protocols	Defines the SSL protocols to be enabled for the SSL context.
-Dzookeeper.ssl.exclude.cipher.ext	Defines the list of passwords separated by a comma which should be excluded from the SSL context.

NOTE

The preceding parameters must be set in the **zkCli.sh/zk.Env.sh** file.

- Server
  a. Set **secureClientPort** to **3381** in the **zoo.cfg** file.
  b. Set **zookeeper.serverCnxnFactory** to **org.apache.zookeeper.server.NettyServerCnxnFactory** in the **zoo.cfg** file on the server.
  c. Set the following parameters in the **zoo.cfg** file (in the **zookeeper/conf/zoo.cfg** path) to configure the server environment:

Parameter	Description
ssl.keyStore.location	Path for storing the **keystore.jks** file
ssl.keyStore.password	Encrypts a password.
ssl.trustStore.location	Indicates the path for storing the truststore file.
ssl.trustStore.password	Encrypts a password.
config.crypt.class	Decrypts an encrypted password.

Parameter	Description
ssl.keyStore.password.encrypted	Default value: **false** If this parameter is set to **true**, the encrypted password can be used.
ssl.trustStore.password.encrypted	Default value: **false** If this parameter is set to **true**, the encrypted password can be used.
ssl.enabled.protocols	Defines the SSL protocols to be enabled for the SSL context.
ssl.exclude.cipher.ext	Defines the list of passwords separated by a comma which should be excluded from the SSL context.

d. Start ZKserver and connect the security client to the security port.

- Credential

  The credential used between client and server in ZooKeeper is **X509AuthenticationProvider**. This credential is initialized using the server certificates specified and trusted by the following parameters:

  – zookeeper.ssl.keyStore.location

  – zookeeper.ssl.keyStore.password

  – zookeeper.ssl.trustStore.location

  – zookeeper.ssl.trustStore.password

  NOTE

  If you do not want to use default mechanism of ZooKeeper, then it can be configured with different trust mechanisms as needed.

1.4 Functions

1.4.1 Multi-tenant

Feature Introduction

Modern enterprises' data clusters are developing towards centralization and cloudification. Enterprise-class big data clusters must meet the following requirements:

- Carry data of different types and formats and run jobs and applications of different types (analysis, query, and stream processing).
- Isolate data of a user from that of another user who has demanding requirements on data security, such as a bank or government institute.

The preceding requirements bring the following challenges to the big data cluster:

- Proper allocation and scheduling of resources to ensure stable operating of applications and jobs
- Strict access control to ensure data and service security

Multi-tenant isolates the resources of a big data cluster into resource sets. Users can lease desired resource sets to run applications and jobs and store data. In a big data cluster, multiple resource sets can be deployed to meet diverse requirements of multiple users.

The MRS big data cluster provides a complete enterprise-class big data multi-tenant solution. Multi-tenant is a collection of multiple resources (each resource set is a tenant) in an MRS big data cluster. It can allocate and schedule resources, including computing and storage resources.

Advantages

- Proper resource configuration and isolation

  The resources of a tenant are isolated from those of another tenant. The resource use of a tenant does not affect other tenants. This mechanism ensures that each tenant can configure resources based on service requirements, improving resource utilization.
- Resource consumption measurement and statistics

  Tenants are system resource applicants and consumers. System resources are planned and allocated based on tenants. Resource consumption by tenants can be measured and recorded.
- Ensured data security and access security

  In multi-tenant scenarios, the data of each tenant is stored separately to ensure data security. The access to tenants' resources is controlled to ensure access security.

Enhanced Schedulers

Schedulers are divided into the open source Capacity scheduler and proprietary Superior scheduler.

To meet enterprise requirements and tackle challenges facing the Yarn community in scheduling, develops the Superior scheduler. In addition to inheriting the advantages of the Capacity scheduler and Fair scheduler, this scheduler is enhanced in the following aspects:

- Enhanced resource sharing policy

  The Superior scheduler supports queue hierarchy. It integrates the functions of open source schedulers and shares resources based on configurable policies. In terms of instances, administrators can use the Superior scheduler to configure an absolute value or percentage policy for queue resources. The resource sharing policy of the Superior scheduler enhances the label scheduling policy of Yarn as a resource pool feature. The nodes in the Yarn cluster can be grouped based on the capacity or service type to ensure that queues can more efficiently utilize resources.
- Tenant-based resource reservation policy

  Resources required by tenants must be ensured for running critical tasks. The Superior scheduler builds a mechanism to support the resource reservation policy. By doing so, reserved resources can be allocated to the tasks run by the tenant queues in a timely manner to ensure proper task execution.

- Fair sharing among tenants and resource pool users

  The Superior scheduler allows shared resources to be configured for users in a queue. Each tenant may have users with different weights. Heavily weighted users may require more shared resources.

- Ensured scheduling performance in a big cluster

  The Superior scheduler receives heartbeats from each NodeManager and saves resource information in memory, which enables the scheduler to control cluster resource usage globally. The Superior scheduler uses the push scheduling model, which makes the scheduling more precise and efficient and remarkably improves cluster resource utilization. Additionally, the Superior scheduler delivers excellent performance when the interval between NodeManager heartbeats is long and prevents heartbeat storms in big clusters.

- Priority policy

  If the minimum resource requirement of a service cannot be met after the service obtains all available resources, a preemption occurs. The preemption function is disabled by default.

1.4.2 Security Hardening

MRS is a platform for massive data management and analysis and has high security. MRS protects user data and service running from the following aspects:

- Network isolation

  The entire system is deployed in a VPC on the public cloud to provide an isolated network environment and ensure service and management security of the cluster. By combining the subnet division, route control, and security group functions of VPC, MRS provides a secure and reliable isolated network environment.

- Resource isolation

  MRS supports resource deployment and isolation of physical resources in dedicated zones. You can flexibly combine computing and storage resources, such as dedicated computing resources + shared storage resources, shared computing resources + dedicated storage resources, and dedicated computing resources + dedicated storage resources.

- Host security

  MRS can be integrated with public cloud security services, including Vulnerability Scan Service (VSS), Host Security Service (HSS), Web Application Firewall (WAF), Cloud Bastion Host (CBH), and Web Tamper Protection (WTP). The following measures are provided to improve security of the OS and ports:

  - Security hardening of OS kernels
  - OS patch update
  - OS permission control
  - OS port management
  - OS protocol and port attack defense

- Application security

  The following measures are used to ensure normal running of big data services:

  - Identification and authentication
  - Web application security
  - Access control
  - Audit security
  - Password security

- Data security

  The following measures are provided to ensure the confidentiality, integrity, and availability of massive amounts of user data:

  - Disaster recovery: MRS supports data backup to OBS and cross-region high reliability.
  - Backup: MRS supports backup of DBService, NameNode, and LDAP metadata and backup of HDFS and HBase service data.

- Data integrity

  Data is verified to ensure its integrity during storage and transmission.

  - CRC32C is used by default to verify the correctness of user data stored in HDFS.
  - DataNodes of HDFS store the verified data. If the data transmitted from a client is abnormal (incomplete), DataNodes report the abnormality to the client, and the client rewrites the data.
  - The client checks data integrity when reading data from a DataNode. If the data is incomplete, the client will read data from another DataNode.

- Data confidentiality

  Based on Apache Hadoop, the distributed file system of MRS supports encrypted storage of files to prevent sensitive data from being stored in plaintext, improving data security. Applications need only to encrypt specified sensitive data. Services are not affected during the encryption process. Based on file system data encryption, Hive provides table-level encryption and HBase provides column family-level encryption. Sensitive data can be encrypted and stored after you specify an encryption algorithm during table creation.

  Encrypted storage and access control of data are used to ensure user data security.

  - HBase stores service data to the HDFS after compression. Users can configure the AES and SMS4 encryption algorithm to encrypt data.
  - All the components allow access permissions to be set for local data directories. Unauthorized users are not allowed to access data.
  - All cluster user information is stored in ciphertext.

- Security authentication

  - Uses a unified user- and role-based authentication system as well as an account- and role-based access control (RBAC) model to centrally control user permissions and batch manage user authorization.

  - Employs Lightweight Directory Access Protocol (LDAP) as an account management system and performs the Kerberos authentication on accounts.
  - Provides the single sign-on (SSO) function that centrally manages and authenticates MRS system and component users.
  - Audits users who have logged in to Manager.

1.4.3 Easy Access to Web UIs of Components

Big data components have their own web UIs to manage their own systems. However, you cannot easily access the web UIs due to network isolation. For example, to access the HDFS web UI, you need to create an ECS to remotely log in to the web UI. This makes the UI access complex and unfriendly.

MRS provides an EIP-based secure channel for you to easily access the web UIs of components. This is more convenient than binding an EIP by yourself, and you can access the web UIs with a few clicks, avoiding the steps for logging in to a VPC, adding security group rules, and obtaining a public IP address. For the Hadoop, Spark, HBase, and Hue components in analysis clusters and the Storm component in streaming clusters, you can quickly access their web UIs from the entries on Manager.

1.4.4 Reliability Enhancement

Based on Apache Hadoop open source software, MRS optimizes and improves the reliability and performance of main service components.

System Reliability

- HA for all management nodes

  In the Hadoop open source version, data and compute nodes are managed in a distributed system, in which a single point of failure (SPOF) does not affect the operation of the entire system. However, a SPOF may occur on management nodes running in centralized mode, which becomes the weakness of the overall system reliability.

  MRS provides similar double-node mechanisms for all management nodes of the service components, such as Manager, HDFS NameNodes, HiveServers, HBase HMasters, Yarn ResourceManagers, KerberosServers, and LdapServers. All of them are deployed in active/standby mode or configured with load sharing, effectively preventing SPOFs from affecting system reliability.

- Reliability guarantee in case of exceptions

  By reliability analysis, the following measures to handle software and hardware exceptions are provided to improve the system reliability:

  - After power supply is restored, services are running properly regardless of a power failure of a single node or the whole cluster, ensuring data reliability in case of unexpected power failures. Key data will not be lost unless the hard disk is damaged.
  - Health status checks and fault handling of the hard disk do not affect services.
  - The file system faults can be automatically handled, and affected services can be automatically restored.

  - The process and node faults can be automatically handled, and affected services can be automatically restored.
  - The network faults can be automatically handled, and affected services can be automatically restored.
- Data backup and restoration

  MRS provides full backup, incremental backup, and restoration functions based on service requirements, preventing the impact of data loss and damages on services and ensuring fast system restoration in case of exceptions.

  - Automatic backup

    MRS provides automatic backup for data on Manager. Based on the customized backup policy, data on clusters, including LdapServer and DBService data, can be automatically backed up.

  - Manual backup

    You can also manually back up data of the cluster management system before the capacity expansion and patch installation to recover the cluster management system functions upon faults.

    To improve the system reliability, data on Manager and HBase is backed up to a third-party server manually.

Node Reliability

- OS health status monitoring

  MRS periodically collects OS hardware resource usage data, including usage of CPUs, memory, hard disks, and network resources.

- Process health status monitoring

  MRS checks the status of service instances and health indicators of service instance processes, enabling you to know the health status of processes in a timely manner.

- Automatic disk troubleshooting

  MRS is enhanced based on the open source version. It can monitor the status of hardware and file systems on all nodes. If an exception occurs, the corresponding partitions will be removed from the storage pool. If a disk is faulty and replaced, a new hard disk will be added for running services. In this case, maintenance operations are simplified. Replacement of faulty disks can be completed online. In addition, users can set hot backup disks to reduce the faulty disk restoration time and improve the system reliability.

- LVM configuration for node disks

  MRS allows you to configure Logic Volume Management (LVM) to plan multiple disks as a logical volume group. Configuring LVM can avoid uneven usage of disks. It is especially important to ensure even usage of disks on components that can use multiple disk capabilities, such as HDFS and Kafka. In addition, LVM supports disk capacity expansion without re-attaching, preventing service interruption.

Data Reliability

MRS can use the anti-affinity node groups and placement group capabilities provided by ECS and the rack awareness capability of Hadoop to redundantly distribute data to multiple physical host machines, preventing data loss caused by physical hardware failures.

1.4.5 Job Management

The job management function provides an entry for you to submit jobs in a cluster, including MapReduce, Spark, HiveQL, and SparkSQL jobs. MRS works with Data Lake Governance Center (DGC) to provide a one-stop big data collaboration development environment and fully-managed big data scheduling capabilities, helping you effortlessly build big data processing centers.

DGC allows you to develop and debug MRS HiveQL/SparkSQL scripts online and develop MRS jobs by performing drag-and-drop operations to migrate and integrate data between MRS and over 20 heterogeneous data sources. Powerful job scheduling and flexible monitoring and alarming help you easily manage data and job O&M.

1.4.6 Bootstrap Actions

Feature Introduction

MRS provides standard elastic big data clusters on the cloud. Nine big data components, such as Hadoop and Spark, can be installed and deployed. Currently, standard cloud big data clusters cannot meet all user requirements, for example, in the following scenarios:

- Common operating system configurations cannot meet data processing requirements, for example, increasing the maximum number of system connections.
- Software tools or running environments need to be installed, for example, Gradle and dependency R language package.
- Big data component packages need to be modified based on service requirements, for example, modifying the Hadoop or Spark installation package.
- Other big data components that are not supported by MRS need to be installed.

To meet the preceding customization requirements, you can manually perform operations on the existing and newly added nodes. The overall process is complex and error-prone. In addition, manual operations cannot be traced, and data cannot be processed immediately after creating a cluster based on your demand.

Therefore, MRS supports custom bootstrap actions that enable you to run scripts on a specified node before or after a cluster component is started. You can run bootstrap actions to install third-party software that is not supported by MRS, modify the cluster running environment, and perform other customizations. If you choose to run bootstrap actions when expanding a cluster, the bootstrap actions will be run on the newly added nodes in the same way. MRS runs the script you specify as user **root**. You can run the **su -** ***xxx*** command in the script to switch the user.

Customer Benefits

You can use the custom bootstrap actions to flexibly and easily configure your dedicated clusters and customize software installation.

1.4.7 Metadata

MRS provides multiple metadata storage methods. When deploying Hive and Ranger during MRS cluster creation, select one of the following storage modes as required:

- **Local**: Metadata is stored in the local GaussDB of a cluster. When the cluster is deleted, the metadata is also deleted. To retain the metadata, manually back up the metadata in the database in advance.
- **Data Connection**: Metadata is stored in the associated PostgreSQL or MySQL database of the RDS service in the same VPC and subnet as the current cluster. When the cluster is terminated, the metadata is not deleted. Multiple MRS clusters can share the metadata.

NOTE

Hive in MRS 1.9.*x* or later allows you to specify a metadata storage method.

Ranger in MRS 1.9.*x* allows metadata to be stored only in the associated MySQL database of the RDS service.

1.4.8 Cluster Management

1.4.8.1 Cluster Lifecycle Management

MRS supports cluster lifecycle management, including creating and terminating clusters.

- Creating a cluster: After you specify a cluster type, components, number of nodes of each type, VM specifications, AZ, VPC, and authentication information, MRS automatically creates a cluster that meets the configuration requirements. You can run customized scripts in the cluster. In addition, you can create clusters of different types for multiple application scenarios, such as Hadoop analysis clusters, HBase clusters, and Kafka clusters. The big data platform supports heterogeneous cluster deployment. That is, VMs of different specifications can be combined in a cluster based on CPU types, disk capacities, disk types, and memory sizes. Various VM specifications can be mixed in a cluster.
- Terminating a cluster: You can terminate a cluster that is no longer needed (including data and configurations in the cluster). MRS will delete all resources related to the cluster.

Creating a Cluster

On the MRS management console, you can create an MRS cluster. You can select a region and cloud resource specifications to create an MRS cluster that is suitable for enterprise services in one click. MRS automatically installs and deploys the enterprise-level big data platform and optimizes parameters based on the selected cluster type, version, and node specifications.

MRS provides you with fully managed big data clusters. When creating a cluster, you can set a VM login mode (password or key pair). You can use all resources of the created MRS cluster. In addition, MRS allows you to deploy a big data cluster on only two ECSs with 4 vCPUs and 8 GB memory, providing more flexible choices for testing and development.

MRS clusters are classified into analysis, streaming, and hybrid clusters.

- Analysis cluster: is used for offline data analysis and provides Hadoop components.
- Streaming cluster: is used for streaming tasks and provides stream processing components.
- Hybrid cluster: is used for not only offline data analysis but also streaming processing, and provides Hadoop components and stream processing components.
- Custom: You can flexibly combine required components (MRS 3.x and later versions) based on service requirements.

MRS cluster nodes are classified into Master, Core, and Task nodes.

- Master node: management node in a cluster. Master processes of a distributed system, Manager, and databases are deployed on Master nodes. Master nodes cannot be scaled out. The processing capability of Master nodes determines the upper limit of the management capability of the entire cluster. MRS supports scale-up of Master node specifications to provide support for management of a larger cluster.
- Core node: used for both storage and computing and can be scaled in or out. Since Core nodes bear data storage, there are many restrictions on scale-in to prevent data loss and auto scaling cannot be performed.
- Task node: used only for computing only and can be scaled in or out. Task nodes bear only computing tasks. Therefore, auto scaling can be performed.

You can create a cluster in two modes: custom create a cluster and quick create a cluster.

- **Custom config**: On the **Custom Config** page, you can flexibly configure cluster parameters based on application scenarios, such as ECS specifications to better suit your service requirements.
- **Quick config**: On the **Quick Config** page, you can quickly create a cluster based on application scenarios, improving cluster configuration efficiency. Currently, Hadoop analysis clusters, HBase clusters, and Kafka clusters are available for your quick creation.
  - Hadoop analysis cluster: uses components in the open-source Hadoop ecosystem to analyze and query vast amounts of data. For example, use Yarn to manage cluster resources, Hive and Spark to provide offline storage and computing of large-scale distributed data, Spark Streaming and Flink to offer streaming data computing, and Presto to enable interactive queries, and Tez to provide a distributed computing framework of directed acyclic graphs (DAGs).
  - HBase cluster: uses Hadoop and HBase components to provide a column-oriented distributed cloud storage system featuring enhanced reliability, excellent performance, and elastic scalability. It applies to the storage and distributed computing of massive amounts of data. You can use HBase to

build a storage system capable of storing TB- or even PB-level data. With HBase, you can filter and analyze data with ease and get responses in milliseconds, rapidly mining data value.

- Kafka cluster: uses Kafka and Storm to provide an open source message system with high throughput and scalability. It is widely used in scenarios such as log collection and monitoring data aggregation to implement efficient streaming data collection and real-time data processing and storage.

Terminating a Cluster

MRS allows you to terminate a cluster when it is no longer needed. After the cluster is terminated, all cloud resources used by the cluster will be released. Before terminating a cluster, you are advised to migrate or back up data. Terminate the cluster only when no service is running in the cluster or the cluster is abnormal and cannot provide services based on O&M analysis. If data is stored on EVS disks or pass-through disks in a big data cluster, the data will be deleted after the cluster is terminated. Therefore, exercise caution when terminating a cluster.

1.4.8.2 Manually Scale Out/In a Cluster

The processing capability of a big data cluster can be horizontally expanded by adding nodes. If the cluster scale does not meet service requirements, you can manually scale out or scale in the cluster. MRS intelligently selects the node with the least load or the minimum amount of data to be migrated for scale-in. The node to be scaled in will not receive new tasks, and continues to execute the existing tasks. At the same time, MRS copies its data to other nodes and the node is decommissioned. If the tasks on the node cannot be completed after a long time, MRS migrates the tasks to other nodes, minimizing the impact on cluster services.

Scaling Out a Cluster

Currently, you can add Core or Task nodes to scale out a cluster to handle peak service loads. The capacity expansion of an MRS cluster node does not affect the services of the existing cluster. For details about data skew caused by capacity expansion, see to rectify the fault.

Scaling In a Cluster

You can reduce the number of Core or Task nodes to scale in a cluster so that MRS delivers better storage and computing capabilities at lower O&M costs based on service requirements. After you scale in an MRS cluster, MRS automatically selects nodes that can be scaled in based on the type of services installed on the nodes.

During the scale-in of Core nodes, data on the original nodes is migrated. If the data location is cached, the client automatically updates the location information, which may affect the latency. Node scale-in may affect the response duration of the first access to some HBase on HDFS data. You can restart HBase or disable or enable related tables to avoid this problem.

Task nodes do not store cluster data. They are compute nodes and do not involve migration of data on the nodes.

1.4.8.3 Auto Scaling

Feature Introduction

More and more enterprises use technologies such as Spark and Hive to analyze data. Processing a large amount of data consumes huge resources and costs much. Typically, enterprises regularly analyze data in a fixed period of time every day rather than all day long. To meet enterprises' requirements, MRS provides the auto scaling function to apply for extra resources during peak hours and release resources during off-peak hours. This enables users to use resources on demand and focus on core business at lower costs.

In big data applications, especially in periodic data analysis and processing scenarios, cluster computing resources need to be dynamically adjusted based on service data changes to meet service requirements. The auto scaling function of MRS enables clusters to be elastically scaled out or in based on cluster loads. In addition, if the data volume changes regularly and you want to scale out or in a cluster before the data volume changes, you can use the MRS resource plan feature.

MRS supports two types of auto scaling policies: auto scaling rules and resource plans

- Auto scaling rules: You can increase or decrease Task nodes based on real-time cluster loads. Auto scaling will be triggered when the data volume changes but there may be some delay.
- Resource plans: If the data volume changes periodically, you can create resource plans to resize the cluster before the data volume changes, thereby avoiding a delay in increasing or decreasing resources.

Both auto scaling rules and resource plans can trigger auto scaling. You can configure both of them or configure one of them. Configuring both resource plans and auto scaling rules improves the cluster node scalability to cope with occasionally unexpected data volume peaks.

In some service scenarios, resources need to be reallocated or service logic needs to be modified after cluster scale-out or scale-in. If you manually scale out or scale in a cluster, you can log in to cluster nodes to reallocate resources or modify service logic. If you use auto scaling, MRS enables you to customize automation scripts for resource reallocation and service logic modification. Automation scripts can be executed before and after auto scaling and automatically adapt to service load changes, all of which eliminates manual operations. In addition, automation scripts can be fully customized and executed at various moments, which can meet your personalized requirements and improve auto scaling flexibility.

Customer Benefits

MRS auto scaling provides the following benefits:

- Reducing costs

  Enterprises do not analyze data all the time but perform a batch data analysis in a specified period of time, for example, 03:00 a.m. The batch analysis may take only two hours.

The auto scaling function enables enterprises to add nodes for batch analysis and automatically releases the nodes after completion of the analysis, minimizing costs.

- Meeting instant query requirements

  Enterprises usually encounter instant analysis tasks, for example, data reports for supporting enterprise decision-making. As a result, resource consumption increases sharply in a short period of time. With the auto scaling function, compute nodes can be added for emergent big data analysis, avoiding a service breakdown due to insufficient compute resources. In this way, you do not need to create extra resources. After the emergency ends, MRS automatically releases the nodes.

- Focusing on core business

  It is difficult for developers to determine resource consumption on the big data secondary development platform because of complex query analysis conditions (such as global sorting, filtering, and merging) and data complexity, for example, uncertainty of incremental data. As a result, estimating the computing volume is difficult. MRS's auto scaling function enable developers to focus on service development without the need for resource estimation.

1.4.8.4 Task Node Creation

Feature Introduction

Task nodes can be created and used for computing only. They do not store persistent data and are the basis for implementing auto scaling.

Customer Benefits

When MRS is used only as a computing resource, Task nodes can be used to reduce costs and facilitate cluster node scaling, flexibly meeting users' requirements for increasing or decreasing cluster computing capabilities.

Application Scenarios

When the data volume change is small in a cluster but the cluster's service processing capabilities need to be remarkably and temporarily improved, add Task nodes to address the following situations:

- The number of temporary services is increased, for example, report processing at the end of the year.
- Long-term tasks need to be completed in a short time, for example, some urgent analysis tasks.

1.4.8.5 Isolating a Host

When detecting that a host is abnormal or faulty and cannot provide services or affects cluster performance, you can exclude the host from the available nodes in the cluster temporarily so that the client can access other available nodes. In scenarios where patches are to be installed in a cluster, you can also exclude a specified node from patch installation. Only non-management nodes can be isolated.

After a host is isolated, all role instances on the host will be stopped, and you cannot start, stop, or configure the host and all instances on the host. In addition, after a host is isolated, statistics about the monitoring status and metric data of hardware and instances on the host cannot be collected or displayed.

1.4.8.6 Managing Tags

Tags are cluster identifiers. Adding tags to clusters can help you identify and manage your cluster resources. By associating with Tag Management Service (TMS), MRS allows users with a large number of cloud resources to tag cloud resources, quickly search for cloud resources with the same tag attribute, and perform unified management operations such as review, modification, and deletion, facilitating unified management of big data clusters and other cloud resources.

You can add a maximum of 10 tags to a cluster when creating the cluster or add them on the details page of the created cluster.

1.4.9 Cluster O&M

Alarm Management

MRS can monitor big data clusters in real time and identify system health status based on alarms and events. In addition, MRS allows you to customize monitoring and alarm thresholds to focus on the health status of each metric. When monitoring data reaches the alarm threshold, the system triggers an alarm.

MRS can also interconnect with the message service system of the Simple Message Notification (SMN) service to push alarm information to users by SMS message or email. For details, see **Message Notification**.

Patch Management

MRS supports cluster patching operations and will release patches for open source big data components in a timely manner. On the MRS cluster management page, you can view patch release information related to running clusters, including the detailed description of the resolved issues and impacts. You can determine whether to install a patch based on the service running status. One-click patch installation involves no manual intervention, and will not cause service interruption through rolling installation, ensuring long-term availability of the clusters.

MRS can display the detailed patch installation process. Patch management also supports patch uninstallation and rollback.

NOTE

MRS 3.x or later does not support patch management on the management console.

O&M Support

Cluster resources provided by MRS belong to users. Generally, when O&M personnel's support is required for troubleshooting of a cluster, O&M personnel cannot directly access the cluster. To better serve customers, MRS provides the following two methods to improve communication efficiency during fault locating:

- Log sharing: You can initiate log sharing on the MRS management console to share a specified log scope with O&M personnel, so that O&M personnel can locate faults without accessing the cluster.
- O&M authorization: If a problem occurs when you use an MRS cluster, you can initiate O&M authorization on the MRS management console. O&M personnel can help you quickly locate the problem, and you can revoke the authorization at any time.

Health Check

MRS provides automatic inspection on system running environments for you to check and audit system running health status in one click, ensuring proper system running and lowering system operation and maintenance costs. After viewing inspection results, you can export reports for archiving and fault analysis.

1.4.10 Message Notification

Feature Introduction

The following operations are often performed during the running of a big data cluster:

- Big data clusters often change, for example, cluster scale-out and scale-in.
- When a service data volume changes abruptly, auto scaling will be triggered.
- After related services are stopped, a big data cluster needs to be stopped.

To immediately notify you of successful operations, cluster unavailability, and node faults, MRS uses Simple Message Notification (SMN) to send notifications to you through SMS and emails, facilitating maintenance.

Customer Benefits

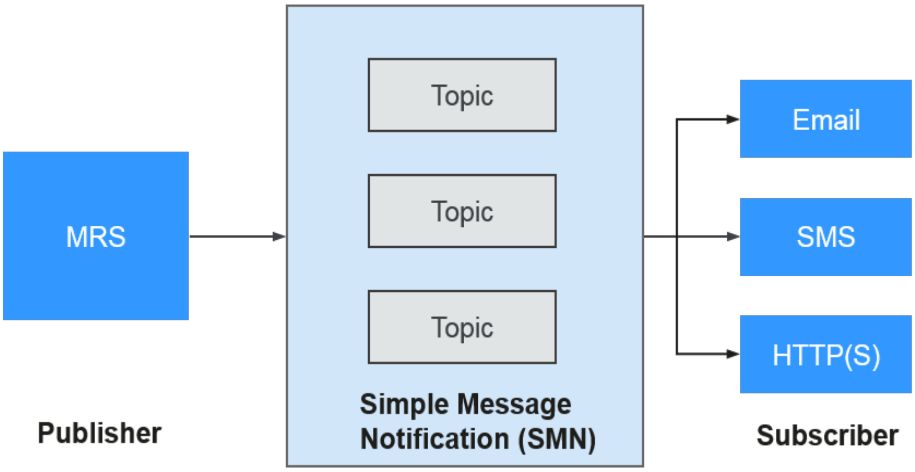
After configuring SMN, you can receive MRS cluster health status, updates, and component alarms through SMS or emails in real time. MRS sends real-time monitoring and alarm notification to help you easily perform O&M and efficiently deploy big data services.

Feature Description

MRS uses SMN to provide one-to-multiple message subscription and notification over a variety of protocols.

You can create a topic and configure topic policies to control publisher and subscriber permissions on the topic. MRS sends cluster messages to the topic to which you have permission to publish messages. Then, all subscribers who subscribe to the topic can receive cluster updates and component alarms through SMS and emails.

Figure 1-143 Implementation process



1.5 Constraints

Before using MRS, ensure that you have read and understand the following restrictions.

- MRS clusters must be created in VPC subnets.
- You are advised to use any of the following browsers to access MRS:
  - Google Chrome: 36.0 or later
  - Internet Explorer: 9.0 or later
- When you create an MRS cluster, you can select **Auto create** from the drop-down list of **Security Group** to create a security group or select an existing security group. After the MRS cluster is created, do not delete or modify the used security group. Otherwise, a cluster exception may occur.
- To prevent illegal access, only assign access permission for security groups used by MRS where necessary.
- Do not perform the following operations because they will cause cluster exceptions:
  - Shutting down, restarting, or deleting MRS cluster nodes displayed in ECS, changing or reinstalling their OS, or modifying their specifications.
  - Deleting the existing processes, applications or files on cluster nodes.
- If a cluster exception occurs when no incorrect operations have been performed, contact technical support engineers. They will ask you for your key and then perform troubleshooting.
- Plan disks of cluster nodes based on service requirements. If you want to store a large volume of service data, add EVS disks or storage space to prevent insufficient storage space from affecting node running.
- The cluster nodes store only users' service data. Non-service data can be stored in the OBS or other ECS nodes.
- The cluster nodes only run MRS cluster programs. Other client applications or user service programs are deployed on separate ECS nodes.
- When you expand the storage capacity of nodes (including master, core, and task) in an MRS cluster, you are advised to create new disks and then attach them to the nodes.

- The capacity (including storage and computing capabilities) of an MRS cluster can be expanded by adding core or task nodes.
- If the cluster is still used to execute tasks or modify configurations after a master node in the cluster is stopped, and other master nodes in the cluster are stopped before the stopped master node is started after task execution or configuration modification, data may be lost due to an active/standby switchover. In this scenario, after the task is executed or the configuration is modified, start the master node that has been stopped and then stop all nodes. If all nodes in the cluster have been stopped, start them in the reverse order of node shutdown.
- The Capacity and Superior scheduler switchover is complete when the MRS cluster is used, while configuration synchronization is not complete. Configure synchronization again based on the new scheduler if necessary.

1.6 Permissions Management

If you need to assign different permissions to employees in your enterprise to access your MRS resources in the cloud, IAM is a good choice for fine-grained permissions management. IAM provides identity authentication, permissions management, and access control, helping you secure access to your cloud resources.

With IAM, you can create IAM users under your cloud account, and assign permissions to these users to control their access to specific resources. For example, some software developers in your enterprise need to use MRS resources but must not delete MRS clusters or perform any high-risk operations. To achieve this goal, you can create IAM users for the software developers and grant them only the permissions required for using MRS cluster resources.

If your cloud account does not require individual IAM users for permissions management, skip this section.

MRS Permission Description

By default, new IAM users do not have any permissions. To assign permissions to a user, add the user to one or more groups and assign permissions policies or roles to these groups. The user then inherits permissions from the groups it is a member of and can perform specified operations on cloud services based on the permissions.

MRS is a project-level service deployed and accessed in specific physical regions. To assign permissions to a user group, specify **Scope** as **Region-specific projects** and select projects in the corresponding region for the permissions to take effect. If **All projects** is selected, the permissions will take effect for the user group in all region-specific projects. When accessing MRS, the users need to switch to a region where they have been authorized to use the MRS service.

You can grant users permissions by using roles and policies.

- Roles: A type of coarse-grained authorization mechanism that defines permissions related to user responsibilities. This mechanism provides only a limited number of service-level roles for authorization. When using roles to grant permissions, you need to also assign other roles on which the permissions depend to take effect. However, roles are not an ideal choice for fine-grained authorization and secure access control.

- Policies: A type of fine-grained authorization mechanism that defines permissions required to perform operations on specific cloud resources under certain conditions. This mechanism allows for more flexible policy-based authorization, meeting requirements for secure access control. For example, you can grant MRS users only the permissions for performing specified operations on MRS clusters, such as creating a cluster and querying a cluster list rather than deleting a cluster. Most policies define permissions based on APIs.

Table 1-29 lists all the system policies supported by MRS.

Table 1-29 MRS system policies

Policy	Description	Type
MRS FullAccess	Administrator permissions for MRS. Users granted these permissions can operate and use all MRS resources.	Fine-grained policy
MRS CommonOperations	Common user permissions for MRS. Users granted these permissions can use MRS but cannot add or delete resources.	Fine-grained policy
MRS ReadOnlyAccess	Read-only permission for MRS. Users granted these permissions can only view MRS resources.	Fine-grained policy
MRS Administrator	Permissions: ● All operations on MRS ● Users with permissions of this policy must also be granted permissions of the **Tenant Guest** and **Server Administrator** policies.	RBAC policy

Table 1-30 lists the common operations supported by each system-defined policy or role of MRS. Select the policies or roles as required.

Table 1-30 Common operations supported by each system-defined policy

Operation	MRS FullAccess	MRS CommonOperations	MRS ReadOnlyAccess	MRS Administrator
Creating a cluster	√	x	x	√
Resizing a cluster	√	x	x	√

Operation	MRS FullAccess	MRS CommonOperations	MRS ReadOnlyAccess	MRS Administrator
Upgrading node specifications	√	x	x	√
Deleting a cluster	√	x	x	√
Querying cluster details	√	√	√	√
Querying a cluster list	√	√	√	√
Configuring an auto scaling rule	√	x	x	√
Querying a host list	√	√	√	√
Querying operation logs	√	√	√	√
Creating and executing a job	√	√	x	√
Stopping a job	√	√	x	√
Deleting a single job	√	√	x	√
Deleting jobs in batches	√	√	x	√
Querying job details	√	√	√	√
Querying a job list	√	√	√	√
Creating a folder	√	√	x	√
Deleting a file	√	√	x	√

Operation	MRS FullAccess	MRS CommonOperations	MRS ReadOnlyAccess	MRS Administrator
Querying a file list	√	√	√	√
Operating cluster tags in batches	√	√	x	√
Creating a single cluster tag	√	√	x	√
Deleting a single cluster tag	√	√	x	√
Querying a resource list by tag	√	√	√	√
Querying cluster tags	√	√	√	√
Accessing Manager	√	√	x	√
Querying a patch list	√	√	√	√
Installing a patch	√	√	x	√
Uninstalling a patch	√	√	x	√
Authorizing O&M channels	√	√	x	√
Sharing O&M channel logs	√	√	x	√
Querying an alarm list	√	√	√	√

Operation	MRS FullAccess	MRS CommonOperations	MRS ReadOnlyAccess	MRS Administrator
Subscribing to alarm notification	√	√	x	√
Submitting an SQL statement	√	√	x	√
Querying SQL results	√	√	x	√
Canceling an SQL execution task	√	√	x	√

MRS FullAccess

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Action": [
                "mrs:*:*",
                "ecs:*:*",
                "bms:*:*",
                "evs:*:*",
                "vpc:*:*",
                "bss:*:*",
                "kms:*:*",
                "rds:*:*"
            ],
            "Effect": "Allow"
        }
    ]
}
```

MRS CommonOperations

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Action": [
                "mrs:*:get*",
                "mrs:*:list*",
                "ecs:*:get*",
                "ecs:*:list*",
                "bms:*:get*",
                "bms:*:list*",
                "evs:*:get*",
                "evs:*:list*",
                "vpc:*:get*",
                "vpc:*:list*",
                "mrs:job:submit",
                "mrs:job:stop",
                "mrs:job:delete",
                "mrs:job:checkSql",
                "mrs:job:batchDelete",
                "mrs:file:create",
                "mrs:file:delete",
                "mrs:tag:batchOperate",
                "mrs:tag:create",
                "mrs:tag:delete",
                "mrs:manager:access",
                "mrs:patch:install",
                "mrs:patch:uninstall",
                "mrs:ops:grant",
                "mrs:ops:shareLog",
                "mrs:alarm:subscribe",
                "mrs:alarm:delete",
                "bss:*:get*",
                "bss:*:list*",
                "kms:*:get*",
                "kms:*:list*",
                "rds:*:get*",
                "rds:*:list*",
                "mrs:bootstrap:*"
            ],
            "Effect": "Allow"
        },
        {
            "Action": [
                "mrs:cluster:create",
                "mrs:cluster:resize",
                "mrs:cluster:scaleUp",
                "mrs:cluster:delete",
                "mrs:cluster:policy"
            ],
            "Effect": "Deny"
        }
    ]
}
```

MRS ReadOnlyAccess

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Action": [
                "mrs:*:get*",
                "mrs:*:list*",
                "mrs:tag:count",
                "ecs:*:get*",
                "ecs:*:list*",
                "bms:*:get*",
                "bms:*:list*",
                "evs:*:get*",
                "evs:*:list*",
                "vpc:*:get*",
                "vpc:*:list*",
                "bss:*:get*",
                "bss:*:list*",
                "kms:*:get*",
                "kms:*:list*",
                "rds:*:get*",
                "rds:*:list*"
            ],
            "Effect": "Allow"
        },
        {
            "Action": [
                "mrs:cluster:create",
                "mrs:cluster:resize",
                "mrs:cluster:scaleUp",
                "mrs:cluster:delete",
                "mrs:cluster:policy",
                "mrs:job:submit",
                "mrs:job:stop",
                "mrs:job:delete",
                "mrs:job:batchDelete",
                "mrs:file:create",
                "mrs:file:delete",
                "mrs:tag:batchOperate",
                "mrs:tag:create",
                "mrs:tag:delete",
                "mrs:manager:access",
                "mrs:patch:install",
                "mrs:patch:uninstall",
                "mrs:ops:grant",
                "mrs:ops:shareLog",
                "mrs:alarm:subscribe"
            ],
            "Effect": "Deny"
        }
    ]
}
```

MRS Administrator

```
{
    "Version": "1.0",
    "Statement": [
        {
            "Action": [
                "MRS:MRS:*"
            ],
            "Effect": "Allow"
        }
    ],
    "Depends": [
        {
            "catalog": "BASE",
            "display_name": "Server Administrator"
        },
        {
            "catalog": "BASE",
            "display_name": "Tenant Guest"
        }
    ]
}
```

1.7 Related Services

Relationships with Other Services

Table 1-31 Relationships with other services

Service	Relationships
Virtual Private Cloud (VPC)	MRS clusters are created in the subnets of a VPC. VPCs provide a secure, isolated, and logical network environment for your MRS clusters.
Object Storage Service (OBS)	OBS stores the following user data: ● MRS job input data, such as user programs and data files ● MRS job output data, such as result files and log files of jobs In MRS clusters, HDFS, Hive, MapReduce, YARN, Spark, Flume, and Loader can import or export data from OBS. MRS uses the parallel file system of OBS to provide services.
Elastic Cloud Server (ECS)	MRS uses elastic cloud servers (ECSs) as cluster nodes.
Relational Database Service (RDS)	RDS stores MRS system running data, including MRS cluster metadata.
Identity and Access Management (IAM)	IAM provides authentication for MRS.
Simple Message Notification (SMN)	MRS uses SMN to provide one-to-multiple message subscription and notification over a variety of protocols.
Cloud Trace Service (CTS)	CTS provides you with operation records of MRS resource operation requests and request results for querying, auditing, and backtracking.

Table 1-32 MRS operations recorded by CTS

Operation	Resource Type	Trace Name
Creating a cluster	cluster_mrs	createCluster
Deleting a cluster	cluster_mrs	deleteCluster
Expanding a cluster	cluster_mrs	scaleOutCluster
Shrinking a cluster	cluster_mrs	scaleInCluster

After you enable CTS, the system starts recording operations on cloud resources. You can view operation records of the last 7 days on the CTS management console. For details, see **Cloud Trace Service** > **Getting Started** > **Querying Real-Time Traces**.

2 Preparing a User

2.1 Creating an MRS User

Use IAM to implement fine-grained permission control over your MRS. With IAM, you can:

- Create IAM users under your account for employees based on your enterprise's organizational structure so that each employee is allowed to access MRS resources using their unique security credential (IAM user).
- Grant only the permissions required for users to perform a specific task.
- Entrust a account or cloud service to perform efficient O&M on your MRS resources.

If your account does not require IAM users, skip this section.

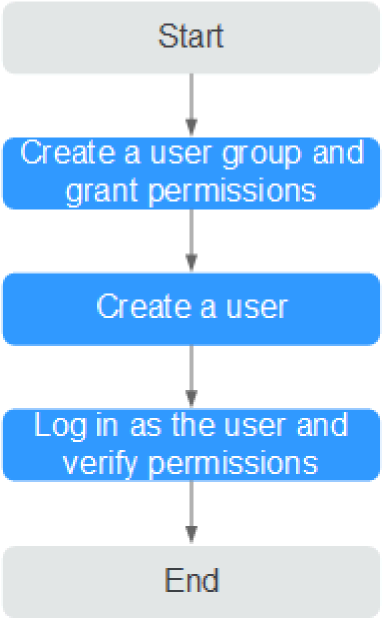
This section describes the procedure for granting permissions (see **Figure 2-1**).

Prerequisites

Learn about the permissions.

Process Flow

Figure 2-1 Process for granting MRS permissions


Start
Create a user group and grant permissions
Create a user
Log in as the user and verify permissions
End

1. **Create a user group and assign permissions to it**.

   Create a user group on the IAM console, and assign MRS permissions to the group.

2. **Create a user and add it to a user group**.

   Create a user on the IAM console and add the user to the group created in **1. Create a user group and assign permissions to it**.

3. Log in and verify permissions.

   Log in to the console by using the user created, and verify that the user has the granted permissions.

   - Choose **Service List** > **MapReduce Service**. Then click **Create Cluster** on the MRS console. If a message appears indicating that you have insufficient permissions to perform the operation, the **MRS ReadOnlyAccess** policy has already taken effect.
   - Choose any other service in **Service List**. If a message appears indicating that you have insufficient permissions to access the service, the **MRS ReadOnlyAccess** policy has already taken effect.

MRS Permission Description

By default, new IAM users do not have any permissions. To assign permissions to a user, add the user to one or more groups and assign permissions policies or roles to these groups. The user then inherits permissions from the groups it is a member of and can perform specified operations on cloud services based on the permissions.

MRS is a project-level service deployed and accessed in specific physical regions. To assign permissions to a user group, specify **Scope** as **Region-specific projects** and select projects in the corresponding region for the permissions to take effect.

If **All projects** is selected, the permissions will take effect for the user group in all region-specific projects. When accessing MRS, the users need to switch to a region where they have been authorized to use the MRS service.

You can grant users permissions by using roles and policies.

- Roles: A type of coarse-grained authorization mechanism that defines permissions related to user responsibilities. This mechanism provides only a limited number of service-level roles for authorization. When using roles to grant permissions, you need to also assign other roles on which the permissions depend to take effect. However, roles are not an ideal choice for fine-grained authorization and secure access control.
- Policies: A type of fine-grained authorization mechanism that defines permissions required to perform operations on specific cloud resources under certain conditions. This mechanism allows for more flexible policy-based authorization, meeting requirements for secure access control. For example, you can grant MRS users only the permissions for performing specified operations on MRS clusters, such as creating a cluster and querying a cluster list rather than deleting a cluster. Most policies define permissions based on APIs.

Table 2-1 lists all the system policies supported by MRS.

Table 2-1 MRS system policies

Policy	Description	Type
MRS FullAccess	Administrator permissions for MRS. Users granted these permissions can operate and use all MRS resources.	Fine-grained policy
MRS CommonOperations	Common user permissions for MRS. Users granted these permissions can use MRS but cannot add or delete resources.	Fine-grained policy
MRS ReadOnlyAccess	Read-only permission for MRS. Users granted these permissions can only view MRS resources.	Fine-grained policy
MRS Administrator	Permissions: ● All operations on MRS ● Users with permissions of this policy must also be granted permissions of the **Tenant Guest** and **Server Administrator** policies.	RBAC policy

Table 2-2 lists the common operations supported by each system-defined policy or role of MRS. Select the policies or roles as required.

Table 2-2 Common operations supported by each system-defined policy

Operation	MRS FullAccess	MRS CommonOperations	MRS ReadOnlyAccess	MRS Administrator
Creating a cluster	√	x	x	√
Resizing a cluster	√	x	x	√
Upgrading node specifications	√	x	x	√
Deleting a cluster	√	x	x	√
Querying cluster details	√	√	√	√
Querying a cluster list	√	√	√	√
Configuring an auto scaling rule	√	x	x	√
Querying a host list	√	√	√	√
Querying operation logs	√	√	√	√
Creating and executing a job	√	√	x	√
Stopping a job	√	√	x	√
Deleting a single job	√	√	x	√
Deleting jobs in batches	√	√	x	√
Querying job details	√	√	√	√

Operation	MRS FullAccess	MRS CommonOperations	MRS ReadOnlyAccess	MRS Administrator
Querying a job list	√	√	√	√
Creating a folder	√	√	x	√
Deleting a file	√	√	x	√
Querying a file list	√	√	√	√
Operating cluster tags in batches	√	√	x	√
Creating a single cluster tag	√	√	x	√
Deleting a single cluster tag	√	√	x	√
Querying a resource list by tag	√	√	√	√
Querying cluster tags	√	√	√	√
Accessing Manager	√	√	x	√
Querying a patch list	√	√	√	√
Installing a patch	√	√	x	√
Uninstalling a patch	√	√	x	√
Authorizing O&M channels	√	√	x	√
Sharing O&M channel logs	√	√	x	√

Operation	MRS FullAccess	MRS CommonOperations	MRS ReadOnlyAccess	MRS Administrator
Querying an alarm list	√	√	√	√
Subscribing to alarm notification	√	√	x	√
Submitting an SQL statement	√	√	x	√
Querying SQL results	√	√	x	√
Canceling an SQL execution task	√	√	x	√

2.2 Creating a Custom Policy

Custom policies can be created to supplement the system-defined policies of MRS. For the actions that can be added to custom policies, see **Permissions Policies and Supported Actions** > **Introduction** in MapReduce Service API Reference.

You can create custom policies in either of the following ways:

- Visual editor: Select cloud services, actions, resources, and request conditions. This does not require knowledge of policy syntax.
- JSON: Edit JSON policies from scratch or based on an existing policy.

Example Custom Policies

- Example 1: Allowing users to create MRS clusters only

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "mrs:cluster:create",
                "ecs:*:*",
                "bms:*:*",
                "evs:*:*",
                "vpc:*:*",
                "smn:*:*"
            ]
        }
    ]
}
```

- Example 2: Allowing users to resize an MRS cluster

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "mrs:cluster:resize"
            ]
        }
    ]
}
```

- Example 3: Allowing users to create a cluster, create and execute a job, and delete a single job, but denying cluster deletion

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "mrs:cluster:create",
                "mrs:job:submit",
                "mrs:job:delete"
            ]
        },
        {
            "Effect": "Deny",
            "Action": [
                "mrs:cluster:delete"
            ]
        }
    ]
}
```

- Example 4: Allowing users to create an ECS cluster with the minimum permission

NOTE

- If you need a key pair when creating a cluster, add the following permissions: **ecs:serverKeypairs:get** and **ecs:serverKeypairs:list**.
- Add the **kms:cmk:list** permission when encrypting data disks during cluster creation.
- Add the **mrs:alarm:subscribe** permission to enable the alarm function during cluster creation.
- Add the **rds:instance:list** permission to use external data sources during cluster creation.

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "mrs:cluster:create"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "ecs:cloudServers:updateMetadata",
                "ecs:cloudServerFlavors:get",
                "ecs:cloudServerQuotas:get",
                "ecs:servers:list",
                "ecs:servers:get",
                "ecs:cloudServers:delete",
                "ecs:cloudServers:list",
                "ecs:serverInterfaces:get",
                "ecs:serverGroups:manage",
                "ecs:servers:setMetadata",
                "ecs:cloudServers:get",
                "ecs:cloudServers:create"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "vpc:securityGroups:create",
                "vpc:securityGroupRules:delete",
                "vpc:vpcs:create",
                "vpc:ports:create",
                "vpc:securityGroups:get",
                "vpc:subnets:create",
                "vpc:privateIps:delete",
                "vpc:quotas:list",
                "vpc:networks:get",
                "vpc:publicIps:list",
                "vpc:securityGroups:delete",
                "vpc:securityGroupRules:create",
                "vpc:privateIps:create",
                "vpc:ports:get",
                "vpc:ports:delete",
                "vpc:publicIps:update",
                "vpc:subnets:get",
                "vpc:publicIps:get",
                "vpc:ports:update",
                "vpc:vpcs:list"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "evs:quotas:get",
                "evs:types:get"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "bms:serverFlavors:get"
            ]
        }
    ]
}
```

- Example 5: Allowing users to create a BMS cluster with the minimum permission

NOTE

- If you need a key pair when creating a cluster, add the following permissions: **ecs:serverKeypairs:get** and **ecs:serverKeypairs:list**.
- Add the **kms:cmk:list** permission when encrypting data disks during cluster creation.
- Add the **mrs:alarm:subscribe** permission to enable the alarm function during cluster creation.
- Add the **rds:instance:list** permission to use external data sources during cluster creation.

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "mrs:cluster:create"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "ecs:servers:list",
                "ecs:servers:get",
                "ecs:cloudServers:delete",
                "ecs:serverInterfaces:get",
                "ecs:serverGroups:manage",
                "ecs:servers:setMetadata",
                "ecs:cloudServers:create",
                "ecs:cloudServerFlavors:get",
                "ecs:cloudServerQuotas:get"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "vpc:securityGroups:create",
                "vpc:securityGroupRules:delete",
                "vpc:vpcs:create",
                "vpc:ports:create",
                "vpc:securityGroups:get",
                "vpc:subnets:create",
                "vpc:privateIps:delete",
                "vpc:quotas:list",
                "vpc:networks:get",
                "vpc:publicIps:list",
                "vpc:securityGroups:delete",
                "vpc:securityGroupRules:create",
                "vpc:privateIps:create",
                "vpc:ports:get",
                "vpc:ports:delete",
                "vpc:publicIps:update",
                "vpc:subnets:get",
                "vpc:publicIps:get",
                "vpc:ports:update",
                "vpc:vpcs:list"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "evs:quotas:get",
                "evs:types:get"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "bms:servers:get",
                "bms:servers:list",
                "bms:serverQuotas:get",
                "bms:servers:updateMetadata",
                "bms:serverFlavors:get"
            ]
        }
    ]
}
```

- Example 6: Allowing users to create a hybrid ECS and BMS cluster with the minimum permission

NOTE

- If you need a key pair when creating a cluster, add the following permissions: **ecs:serverKeypairs:get** and **ecs:serverKeypairs:list**.
- Add the **kms:cmk:list** permission when encrypting data disks during cluster creation.
- Add the **mrs:alarm:subscribe** permission to enable the alarm function during cluster creation.
- Add the **rds:instance:list** permission to use external data sources during cluster creation.

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "mrs:cluster:create"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "ecs:cloudServers:updateMetadata",
                "ecs:cloudServerFlavors:get",
                "ecs:cloudServerQuotas:get",
                "ecs:servers:list",
                "ecs:servers:get",
                "ecs:cloudServers:delete",
                "ecs:cloudServers:list",
                "ecs:serverInterfaces:get",
                "ecs:serverGroups:manage",
                "ecs:servers:setMetadata",
                "ecs:cloudServers:get",
                "ecs:cloudServers:create"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "vpc:securityGroups:create",
                "vpc:securityGroupRules:delete",
                "vpc:vpcs:create",
                "vpc:ports:create",
                "vpc:securityGroups:get",
                "vpc:subnets:create",
                "vpc:privateIps:delete",
                "vpc:quotas:list",
                "vpc:networks:get",
                "vpc:publicIps:list",
                "vpc:securityGroups:delete",
                "vpc:securityGroupRules:create",
                "vpc:privateIps:create",
                "vpc:ports:get",
                "vpc:ports:delete",
                "vpc:publicIps:update",
                "vpc:subnets:get",
                "vpc:publicIps:get",
                "vpc:ports:update",
                "vpc:vpcs:list"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "evs:quotas:get",
                "evs:types:get"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "bms:servers:get",
                "bms:servers:list",
                "bms:serverQuotas:get",
                "bms:servers:updateMetadata",
                "bms:serverFlavors:get"
            ]
        }
    ]
}
```

2.3 Synchronizing IAM Users to MRS

IAM user synchronization is to synchronize IAM users bound with MRS policies to the MRS system and create accounts with the same usernames but different passwords as the IAM users. Then, you can use an IAM username (the password needs to be reset by user **admin** of Manager) to log in to Manager for cluster management, and submit jobs on the GUI in a cluster with Kerberos authentication enabled.

Table 2-3 compares IAM users' permission policies and the synchronized users' permissions on MRS. For details about the default permissions on Manager, see **Users and Permissions of MRS Clusters**.

Table 2-3 Policy and permission mapping after synchronization

Policy Type	IAM Policy	User's Default Permissions on MRS After Synchronization	Have Permission to Perform the Synchronization	Have Permission to Submit Jobs
Fine-grained	MRS ReadOnlyAccess	Manager_viewer	No	No
	MRS CommonOperations	● Manager_viewer ● default ● launcher-job	No	Yes

Policy Type	IAM Policy	User's Default Permissions on MRS After Synchronization	Have Permission to Perform the Synchronization	Have Permission to Submit Jobs
	MRS FullAccess	● Manager_administrator ● Manager_auditor ● Manager_operator ● Manager_tenant ● Manager_viewer ● System_administrator ● default ● launcher-job	Yes	Yes
RBAC	MRS Administrator	● Manager_administrator ● Manager_auditor ● Manager_operator ● Manager_tenant ● Manager_viewer ● System_administrator ● default ● launcher-job	No	Yes

Policy Type	IAM Policy	User's Default Permissions on MRS After Synchronization	Have Permission to Perform the Synchronization	Have Permission to Submit Jobs
	Server Administrator, Tenant Guest, and MRS Administrator	● Manager_administrator ● Manager_auditor ● Manager_operator ● Manager_tenant ● Manager_viewer ● System_administrator ● default ● launcher-job	Yes	Yes
	Tenant Administrator	● Manager_administrator ● Manager_auditor ● Manager_operator ● Manager_tenant ● Manager_viewer ● System_administrator ● default ● launcher-job	Yes	Yes

Policy Type	IAM Policy	User's Default Permissions on MRS After Synchronization	Have Permission to Perform the Synchronization	Have Permission to Submit Jobs
Custom	Custom policy	● Manager_viewer ● default ● launcher-job	● If custom policies use RBAC policies as a template, refer to the RBAC policies. ● If custom policies use fine-grained policies as a template, refer to the fine-grained policies. The fine-grained policies are recommended.	Yes

NOTE

To facilitate user permission management, use fine-grained policies rather than RBAC policies. In fine-grained policies, the Deny action takes precedence over other actions.

- A user has permission to synchronize IAM users only when the user has the Tenant Administrator role or has the Server Administrator, Tenant Guest, and MRS Administrator roles at the same time.
- A user with the **action:mrs:cluster:syncUser** policy has permission to synchronize IAM users.

Procedure

Step 1 Create a user and authorize the user to use MRS. For details, see **Creating an MRS User**.

Step 2 Log in to the MRS management console and create a cluster. For details, see **Creating a Custom Cluster**.

Step 3 In the left navigation pane, choose **Clusters** > **Active Clusters**. Click the cluster name to go to the cluster details page.

Step 4 In the **Basic Information** area on the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.

Step 5 After a synchronization request is sent, choose **Operation Logs** in the left navigation pane on the MRS console to check whether the synchronization is successful. For details about the logs, see **Viewing MRS Operation Logs**.

Step 6 After the synchronization is successful, use the user synchronized with IAM to perform subsequent operations.

NOTE

- When the policy of the user group to which the IAM user belongs changes from **MRS ReadOnlyAccess** to **MRS CommonOperations**, **MRS FullAccess**, or **MRS Administrator**, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** (System Security Services Daemon) cache of cluster nodes needs time to be updated. Then, submit a job. Otherwise, the job may fail to be submitted.
- When the policy of the user group to which the IAM user belongs changes from **MRS CommonOperations**, **MRS FullAccess**, or **MRS Administrator** to **MRS ReadOnlyAccess**, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** cache of cluster nodes needs time to be updated.
- After you click **Synchronize** on the right side of **IAM User Sync**, the cluster details page is blank for a short time, because user data is being synchronized. The page will be properly displayed after the data synchronization is complete.

- Submitting jobs in a security cluster: Users can submit jobs using the job management function on the GUI in the security cluster. For details, see **Running a MapReduce Job**.
- All tabs are displayed on the cluster details page, including **Components**, **Tenants**, and **Backups & Restorations**.
- Logging in to Manager
  a. Log in to Manager as user **admin**. For details, see **Accessing Manager**.
  b. Initialize the password of the user synchronized with IAM. For details, see **Initializing the Password of a System User**.
  c. Modify the role bound to the user group to which the user belongs to control user permissions on Manager. For details, see **Related Tasks**. For details about how to create and modify a role, see **Creating a Role**. After the component role bound to the user group to which the user belongs is modified, it takes some time for the role permissions to take effect.
  d. Log in to Manager using the user synchronized with IAM and the password after the initialization in **Step 6.b**.

NOTE

If the IAM user's permission changes, go to **Step 4** to perform second synchronization. After the second synchronization, a system user's permissions are the union of the permissions defined in the IAM system policy and the permissions of roles added by the system user on Manager. After the second synchronization, a custom user's permissions are subject to the permissions configured on Manager.

- System user: If all user groups to which an IAM user belongs are bound to system policies (RABC policies and fine-grained policies belong to system policies), the IAM user is a system user.
- Custom user: If the user group to which an IAM user belongs is bound to any custom policy, the IAM user is a custom user.

----End

3 Configuring a Cluster

3.1 Methods of Creating MRS Clusters

This section describes how to create MRS clusters.

- **Quick Creation of a Hadoop Analysis Cluster**: On the **Quick Config** tab page, you can quickly configure parameters to create Hadoop analysis clusters within a few minutes, facilitating analysis and queries of vast amounts of data.
- **Quick Creation of an HBase Analysis Cluster**: On the **Quick Config** tab page, you can quickly configure parameters to create HBase query clusters within a few minutes, facilitating storage and distributed computing of vast amounts of data.
- **Quick Creation of a Kafka Streaming Cluster**: On the **Quick Config** tab page, you can quickly configure parameters to create Kafka streaming clusters within a few minutes, facilitating streaming data ingestion as well as real-time data processing and storage.
- **Quick Creation of a ClickHouse Cluster**: You can quickly create a ClickHouse cluster. ClickHouse is a columnar database management system used for online analysis. It features the ultimate compression rate and fast query performance.
- **Quick Creation of a Real-time Analysis Cluster**: You can create a real-time analysis cluster within a few minutes to quickly collect, analyze, and query a large amount of data.
- **Creating a Custom Cluster**: On the **Custom Config** tab page, you can flexibly configure parameters to create clusters based on application scenarios, such as ECS specifications to better suit your service requirements.

3.2 Quick Creation of a Cluster

3.2.1 Quick Creation of a Hadoop Analysis Cluster

This section describes how to quickly create a Hadoop analysis cluster for analyzing and querying vast amounts of data. In the open-source Hadoop ecosystem, Hadoop uses Yarn to manage cluster resources, Hive and Spark to provide offline storage and computing of large-scale distributed data, Spark Streaming and Flink to offer streaming data computing, and Presto to enable interactive queries, Tez to provide a distributed computing framework of directed acyclic graphs (DAGs).

Quick Creation of a Hadoop Analysis Cluster

Step 1 Log in to the MRS console.

Step 2 Click **Create Cluster**. The page for creating a cluster is displayed.

Step 3 Click the **Quick Config** tab.

Step 4 Configure basic cluster information. For details about the parameters, see **Creating a Custom Cluster**.

- **Region**: Use the default value.
- **Cluster Name**: You can use the default name. However, you are advised to include a project name abbreviation or date for consolidated memory and easy distinguishing, for example, **mrs_20180321**.
- **Cluster Version**: Select the latest version, which is the default value. (The components provided by a cluster vary according to the cluster version. Select a cluster version based on site requirements.)
- **Component**: Select **Hadoop analysis cluster**.
- **AZ**: Use the default value.
- **VPC**: Use the default value. If there is no available VPC, click **View VPC** to access the VPC console and create a new VPC.
- **Subnet**: Use the default value.
- **Enterprise Project**: Use the default value.
- **Cluster Node**: Select the number of cluster nodes and node specifications based on site requirements. For MRS 3.*x* or later, the memory of the master node must be greater than 64 GB.
- **Cluster HA**: Use the default value. This parameter is not available in MRS 3.*x*.
- **Kerberos Authentication**: Select whether to enable Kerberos authentication.
- **Username**: The default username is **admin**, which is used to log in to MRS Manager.
- **Password**: Set a password for user **admin**.
- **Confirm Password**: Enter the password of user **admin** again.
- **Key Pair**: Select a key pair from the drop-down list to log in to an ECS. Select **"I acknowledge that I have obtained private key file *SSHkey-xxx* and that without this file I will not be able to log in to my ECS."** If you have never created a key pair, click **View Key Pair** to create or import a key pair. And then, obtain a private key file.

Step 5 Select **Enable** to enable secure communications. For details, see **Communication Security Authorization**.

Step 6 Click **Create Now**.

If Kerberos authentication is enabled for a cluster, check whether Kerberos authentication is required. If yes, click **Continue**. If no, click **Back** to disable Kerberos authentication and then create a cluster.

Step 7 Click **Back to Cluster List** to view the cluster status. Click **Access Cluster** to view cluster details.

For details about cluster status during creation, see the description of the status parameters in **Table 4-4**.

It takes some time to create a cluster. The initial status of the cluster is **Starting**. After the cluster has been created successfully, the cluster status becomes **Running**.

On the MRS management console, a maximum of 10 clusters can be concurrently created, and a maximum of 100 clusters can be managed.

----End

3.2.2 Quick Creation of an HBase Analysis Cluster

This section describes how to quickly create an HBase query cluster. The HBase cluster uses Hadoop and HBase components to provide a column-oriented distributed cloud storage system featuring enhanced reliability, excellent performance, and elastic scalability. It applies to the storage and distributed computing of massive amounts of data. You can use HBase to build a storage system capable of storing TB- or even PB-level data. With HBase, you can filter and analyze data with ease and get responses in milliseconds, rapidly mining data value.

Quick Creation of an HBase Analysis Cluster

Step 1 Log in to the MRS console.

Step 2 Click **Create Cluster**. The page for creating a cluster is displayed.

Step 3 Click the **Quick Config** tab.

Step 4 Configure basic cluster information. For details about the parameters, see **Creating a Custom Cluster**.

- **Region**: Use the default value.
- **Cluster Name**: You can use the default name. However, you are advised to include a project name abbreviation or date for consolidated memory and easy distinguishing, for example, **mrs_20180321**.
- **Cluster Version**: Select the latest version, which is the default value. (The components provided by a cluster vary according to the cluster version. Select a cluster version based on site requirements.)
- **Component**: Select **HBase Query Cluster**.
- **AZ**: Use the default value.
- **VPC**: Use the default value. If there is no available VPC, click **View VPC** to access the VPC console and create a new VPC.
- **Subnet**: Use the default value.
- **Enterprise Project**: Use the default value.

- **Cluster Node**: Select the number of cluster nodes and node specifications based on site requirements. For MRS 3.*x* or later, the memory of the master node must be greater than 64 GB.
- **Cluster HA**: Use the default value. This parameter is not available in MRS 3.*x*.
- **Kerberos Authentication**: Select whether to enable Kerberos authentication.
- **Username**: The default username is **admin**, which is used to log in to MRS Manager.
- **Password**: Set a password for user **admin**.
- **Confirm Password**: Enter the password of user **admin** again.
- **Key Pair**: Select a key pair from the drop-down list to log in to an ECS. Select **"I acknowledge that I have obtained private key file** *SSHkey-xxx* **and that without this file I will not be able to log in to my ECS.**" If you have never created a key pair, click **View Key Pair** to create or import a key pair. And then, obtain a private key file.

Step 5 Select **Enable** to enable secure communications. For details, see **Communication Security Authorization**.

Step 6 Click **Create Now**.

If Kerberos authentication is enabled for a cluster, check whether Kerberos authentication is required. If yes, click **Continue**. If no, click **Back** to disable Kerberos authentication and then create a cluster.

Step 7 Click **Back to Cluster List** to view the cluster status. Click **Access Cluster** to view cluster details.

For details about cluster status during creation, see the description of the status parameters in **Table 4-4**.

It takes some time to create a cluster. The initial status of the cluster is **Starting**. After the cluster has been created successfully, the cluster status becomes **Running**.

On the MRS management console, a maximum of 10 clusters can be concurrently created, and a maximum of 100 clusters can be managed.

----End

3.2.3 Quick Creation of a Kafka Streaming Cluster

This section describes how to quickly create a Kafka streaming cluster. The Kafka cluster uses the Kafka and Storm components to provide an open-source messaging system with high throughput and scalability. It is widely used in scenarios such as log collection and monitoring data aggregation to implement efficient streaming data collection and real-time data processing and storage.

Quick Creation of a Kafka Streaming Cluster

Step 1 Log in to the MRS console.

Step 2 Click **Create Cluster**. The page for creating a cluster is displayed.

Step 3 Click the **Quick Config** tab.

Step 4 Configure basic cluster information. For details about the parameters, see **Creating a Custom Cluster**.

- **Region**: Use the default value.
- **Cluster Name**: You can use the default name. However, you are advised to include a project name abbreviation or date for consolidated memory and easy distinguishing, for example, **mrs_20200321**.
- **Cluster Version**: The components provided by a cluster vary according to the cluster version. Select a cluster version based on site requirements.
- **Component**: Select **Kafka streaming cluster**.
- **AZ**: Use the default value.
- **VPC**: Use the default value. If there is no available VPC, click **View VPC** to access the VPC console and create a new VPC.
- **Subnet**: Use the default value.
- **Cluster Node**: Select the number of cluster nodes and node specifications based on site requirements. For MRS 3.*x* or later, the memory of the master node must be greater than 64 GB.
- **Cluster HA**: Use the default value. This parameter is not available in MRS 3.*x*.
- **Username**: The default username is **admin**, which is used to log in to MRS Manager.
- **Password**: Set a password for user **admin**.
- **Confirm Password**: Enter the password of user **admin** again.
- **Key Pair**: Select a key pair from the drop-down list to log in to an ECS. Select **"I acknowledge that I have obtained private key file *SSHkey-xxx* and that without this file I will not be able to log in to my ECS.**" If you have never created a key pair, click **View Key Pair** to create or import a key pair. And then, obtain a private key file.

Step 5 Select **Enable** to enable secure communications. For details, see **Communication Security Authorization**.

Step 6 **Click Apply Now.**

If Kerberos authentication is enabled for a cluster, check whether Kerberos authentication is required. If yes, click **Continue**. If no, click **Back** to disable Kerberos authentication and then create a cluster.

Step 7 Click **Back to Cluster List** to view the cluster status. Click **Access Cluster** to view cluster details.

For details about cluster status during creation, see the description of the status parameters in **Table 4-4**.

It takes some time to create a cluster. The initial status of the cluster is **Starting**. After the cluster has been created successfully, the cluster status becomes **Running**.

On the MRS management console, a maximum of 10 clusters can be concurrently created, and a maximum of 100 clusters can be managed.

----End

3.2.4 Quick Creation of a ClickHouse Cluster

This section describes how to quickly create a ClickHouse cluster. ClickHouse is a columnar database management system used for online analysis. It features the ultimate compression rate and fast query performance. It is widely used in Internet advertisement, app and web traffic analysis, telecom, finance, and IoT fields.

The ClickHouse cluster consists of the following components:

- MRS 3.1.0: ClickHouse 21.3.4.25 and ZooKeeper 3.5.6.

The ClickHouse cluster table engine that uses Kunpeng as the CPU architecture does not support HDFS and Kafka.

Quick Creation of a ClickHouse Cluster

Step 1 Log in to the MRS console.

Step 2 Click **Create Cluster**. The page for creating a cluster is displayed.

Step 3 Click the **Quick Config** tab.

Step 4 Configure basic cluster information. For details about the parameters, see **Creating a Custom Cluster**.

- **Region**: Use the default value.
- **Cluster Name**: You can use the default name. However, you are advised to include a project name abbreviation or date for consolidated memory and easy distinguishing, Example: **mrs_20201121**.
- **Cluster Version**: Select the latest version, which is the default value. (The components provided by a cluster vary according to the cluster version. Select a cluster version based on site requirements.)
- **Component**: Select **ClickHouse cluster**.
- **AZ**: Use the default value.
- **VPC**: Use the default value. If there is no available VPC, click **View VPC** to access the VPC console and create a new VPC.
- **Subnet**: Use the default value.
- **Cluster Node**: Select the number of cluster nodes and node specifications based on site requirements. For MRS 3.*x* or later, the memory of the master node must be greater than 64 GB.
- **Username**: The default value is **root/admin**. User **root** is used to remotely log in to ECSs, and user **admin** is used to access the cluster management page.

Step 5 Select **Enable** to enable secure communications. For details, see **Communication Security Authorization**.

Step 6 **Click Apply Now.**

If Kerberos authentication is enabled for a cluster, check whether Kerberos authentication is required. If yes, click **Continue**. If no, click **Back** to disable Kerberos authentication and then create a cluster.

Step 7 Click **Back to Cluster List** to view the cluster status. Click **Access Cluster** to view cluster details.

For details about cluster status during creation, see the description of the status parameters in **Table 4-4**.

It takes some time to create a cluster. The initial status of the cluster is **Starting**. After the cluster has been created successfully, the cluster status becomes **Running**.

On the MRS management console, a maximum of 10 clusters can be concurrently created, and a maximum of 100 clusters can be managed.

----End

3.2.5 Quick Creation of a Real-time Analysis Cluster

This section describes how to quickly create a real-time analysis cluster. The real-time analysis cluster uses Hadoop, Kafka, Flink, and ClickHouse to collect, analyze, and query a large amount of data in real time.

The real-time analysis cluster consists of the following components:

- MRS 3.1.0: Hadoop 3.1.1, Kafka 2.11-2.4.0, Flink 1.12.0, ClickHouse 21.3.4.25, ZooKeeper 3.5.6, and Ranger 2.0.0.
- MRS 3.1.2-LTS.3: Hadoop 3.1.1, Kafka 2.11-2.4.0, Flink 1.12.2, ClickHouse 21.3.4.25, ZooKeeper 3.6.3, and Ranger 2.0.0.

Quick Creation of a Real-time Analysis Cluster

Step 1 Log in to the MRS console.

Step 2 Click **Create Cluster**. The page for creating a cluster is displayed.

Step 3 Click the **Quick Config** tab.

Step 4 Configure basic cluster information. For details about the parameters, see **Creating a Custom Cluster**.

- **Region**: Use the default value.
- **Cluster Name**: You can use the default name. However, you are advised to include a project name abbreviation or date for consolidated memory and easy distinguishing, Example: **mrs_20201130**.
- **Cluster Version**: Select the latest version, which is the default value. (The components provided by a cluster vary according to the cluster version. Select a cluster version based on site requirements.)
- **Component**: Select **Real-time Analysis Cluster**.
- **AZ**: Use the default value.
- **VPC**: Use the default value. If there is no available VPC, click **View VPC** to access the VPC console and create a new VPC.
- **Subnet**: Use the default value.
- **Enterprise Project**: Use the default value.
- **Cluster Node**: Select the number of cluster nodes and node specifications based on site requirements. For MRS 3.*x* or later, the memory of the master node must be greater than 64 GB.
- **Username**: The default username is **admin**, which is used to log in to FusionInsight Manager.

- **Password**: Set a password for user **admin**.
- **Confirm Password**: Enter the password of user **admin** again.
- **Key Pair**: Select a key pair from the drop-down list to log in to an ECS. Select **I acknowledge that I have obtained private key file** ***SSHkey-xxx*** **and that without this file I will not be able to log in to my ECS.**. If you have not created any key pairs, click **View Key Pair** to create or import one. Then, obtain a private key file.

Step 5 Select **Enable** to enable secure communications. For details, see **Communication Security Authorization**.

Step 6 Click **Create Now**.

If Kerberos authentication is enabled for a cluster, check whether Kerberos authentication is required. If yes, click **Continue**. If no, click **Back** to disable Kerberos authentication and then create a cluster.

Step 7 Click **Back to Cluster List** to view the cluster status. Click **Access Cluster** to view cluster details.

For details about cluster status during creation, see the description of the status parameters in **Table 4-4**.

It takes some time to create a cluster. The initial status of the cluster is **Starting**. After the cluster has been created successfully, the cluster status becomes **Running**.

On the MRS management console, a maximum of 10 clusters can be concurrently created, and a maximum of 100 clusters can be managed.

----End

3.3 Creating a Custom Cluster

The first step of using MRS is to create a cluster. This section describes how to create a cluster on the **Custom Config** tab of the MRS management console.

You can create an IAM user or user group on the IAM management console and grant it specific operation permissions, to perform refined resource management after registering an account. For details, see **Creating an MRS User**.

Step 1 Log in to the MRS console.

Step 2 Click **Create Cluster**. The page for creating a cluster is displayed.

NOTE

When creating a cluster, pay attention to quota notification. If a resource quota is insufficient, increase the resource quota as prompted and create a cluster.

Step 3 Click the **Custom Config** tab.

Step 4 Configure cluster information by referring to **Software Configurations** and click **Next**.

Step 5 Configure cluster information by referring to **Hardware Configurations** and click **Next**.

Step 6 Set advanced options by referring to **(Optional) Advanced Configuration** and click **Create Now**.

If Kerberos authentication is enabled for a cluster, check whether Kerberos authentication is required. If yes, click **Continue**. If no, click **Back** to disable Kerberos authentication and then create a cluster.

Step 7 Click **Back to Cluster List** to view the cluster status.

For details about cluster status during creation, see the description of the status parameters in **Table 4-4**.

It takes some time to create a cluster. The initial status of the cluster is **Starting**. After the cluster has been created successfully, the cluster status becomes **Running**.

On the MRS management console, a maximum of 10 clusters can be concurrently created, and a maximum of 100 clusters can be managed.

----End

Software Configurations

Table 3-1 MRS cluster software configuration

Parameter	Description
Region	Select a region. Cloud service products in different regions cannot communicate with each other over an intranet. For low network latency and quick access, select the nearest region.
Cluster Name	The cluster name must be unique. A cluster name can contain 1 to 64 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed. The default name is **mrs_***xxxx*. *xxxx* is a random collection of letters and digits.
Cluster Version	Currently, MRS 1.8.9, 2.0.1, 3.1.0-LTS.1, and 3.1.2-LTS.3 are supported.

Parameter	Description
Cluster Type	The cluster types are as follows: ● **Analysis cluster**: is used for offline data analysis and provides Hadoop components. ● **Streaming cluster**: is used for streaming tasks and provides stream processing components. ● **Hybrid cluster**: is used for both offline data analysis and streaming processing and provides Hadoop components and streaming processing components. You are advised to use a hybrid cluster to perform offline data analysis and streaming processing tasks at the same time. ● **Custom**: You can adjust the cluster service deployment mode based on service requirements. For details, see **Creating a Custom Topology Cluster**. (This type is currently available only in MRS 3.*x*.) **NOTE** ● MRS streaming clusters do not support job and file management functions. ● To install all components in a cluster, select **Custom**.

Parameter	Description
Components	MRS components are as follows.. **Components of an analysis cluster:** ● Presto: open source and distributed SQL query engine ● Hadoop: distributed system architecture ● Spark: in-memory distributed computing framework (not supported in MRS 3.*x*) ● Spark2x: A fast general-purpose engine for large-scale data processing. It is developed based on the open-source Spark2.x version. (supported only in MRS 3.*x*) ● Hive: data warehouse framework built on Hadoop ● OpenTSDB: a distributed, scalable time series database that can store and serve massive amounts of time series data without losing granularity (not supported in MRS 3.*x*) ● HBase: distributed column-oriented database ● Tez: an application framework which allows for a complex directed-acyclic-graph of tasks for processing data ● Hue: provides the Hadoop UI capability, which enables users to analyze and process Hadoop cluster data on browsers ● Loader: a tool based on source Sqoop 1.99.7, designed for efficiently transferring bulk data between Apache Hadoop and structured datastores such as relational databases (not supported in MRS 3.*x*) Hadoop is mandatory, and Spark and Hive must be used together. Select components based on service requirements. ● Flink: a distributed big data processing engine that can perform stateful computations over both finite and infinite data streams ● Oozie: a Hadoop job scheduling system (supported only in MRS 3.*x*) ● HetuEngine: a distributed SQL query engine for heterogeneous big data sets (supported only in MRS 3.1.*x*-LTS) ● Ranger: a framework to enable, monitor, and manage data security across the Hadoop platform ● Impala: an SQL query engine for processing huge volumes of data ● ClickHouse: A column database management system (DBMS) for on-line analytical processing

Parameter	Description
	(OLAP). The ClickHouse cluster table engine that uses Kunpeng as the CPU architecture does not support HDFS and Kafka. ● Kudu: a column-oriented data store **Components of a streaming cluster:** ● Kafka: distributed messaging system ● Flume: distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data ● ZooKeeper: a centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services (supported only in MRS 3.*x*) ● Ranger: a framework to enable, monitor, and manage data security across the Hadoop platform (supported only in MRS 3.*x*)
Metadata	Whether to use external data sources to store metadata. ● **Local**: Metadata is stored in the local cluster. ● **Data connections**: Metadata of external data sources is used. If the cluster is abnormal or deleted, metadata is not affected. This mode applies to scenarios where storage and compute are decoupled. Clusters that support the Hive or Ranger component support this function.
Component	This parameter is valid only when **Metadata** is set to **Data connections**. It indicates the type of an external data source. ● Hive ● Ranger
Data Connection Type	This parameter is valid only when **Metadata** is set to **Data connections**. It indicates the type of an external data source. ● Hive supports the following data connection types: – RDS MySQL database – Local database ● Ranger supports the following data connection types: – RDS MySQL database – Local database

Parameter	Description
Data Connection Instance	This parameter is valid only when **Data Connection Type** is set to **RDS MySQL database**. This parameter indicates the name of the connection between the MRS cluster and the RDS database. This instance must be created before being referenced here. You can click **Create Data Connection** to create a data connection. For details, see **Configuring Data Connections**.

Hardware Configurations

Table 3-2 MRS cluster hardware configuration

Parameter	Description
AZ	Select the AZ associated with the region of the cluster. An AZ is a physical area that uses independent power and network resources. AZs are physically isolated but interconnected through the internal network. This improves the availability of applications. You are advised to create clusters in different AZs.
VPC	A VPC is a secure, isolated, and logical network environment. Select the VPC for which you want to create a cluster and click **View VPC** to view the name and ID of the VPC. If no VPC is available, create one.

Parameter	Description
Subnet	A subnet provides dedicated network resources that are isolated from other networks, improving network security. Select the subnet for which you want to create a cluster. Click **View Subnet** to view details about the selected subnet. If no subnet is created in the VPC, go to the VPC console and choose **Subnets** > **Create Subnet** to create one. For details about how to configure network ACL outbound rules, see **How Do I Configure a Network ACL Outbound Rule?** **NOTE** The number of IP addresses required by creating an MRS cluster depends on the number of cluster nodes and selected components, but not the cluster type. In MRS, IP addresses are automatically assigned to clusters during cluster creation basically based on the following formula: Quantity of IP addresses = Number of cluster nodes + 2 (one for Manager; one for the DB). In addition, if the Hadoop, Hue, Sqoop, and Presto or Loader and Presto components are selected during cluster deployment, one IP address is added for each component. To create a ClickHouse cluster independently, the number of IP addresses required is calculated as follows: Number of IP addresses = Number of cluster nodes + 1 (for Manager).
Security Group	A security group is a set of ECS access rules. It provides access policies for ECSs that have the same security protection requirements and are mutually trusted in a VPC. When you create a cluster, you can select **Auto create** from the drop-down list of **Security Group** to create a security group or select an existing security group. **NOTE** When you select a security group created by yourself, ensure that the inbound rule contains a rule in which **Protocol** is set to **All**, **Port** is set to **All**, and **Source** is set to a trusted accessible IP address range. Do not use **0.0.0.0/0** as a source address. Otherwise, security risks may occur. If you do not know the trusted accessible IP address range, select **Auto create**.
EIP	After binding an EIP to an MRS cluster, you can use the EIP to access the Manager web UI of the cluster. When creating a cluster, you can select an available EIP from the drop-down list and bind it. If no EIP is available in the drop-down list, click **Manage EIP** to access the **EIPs** service page to create one. **NOTE** The EIP must be in the same region as the cluster.

Table 3-3 Cluster node information

Parameter	Description
Common Template	This parameter is valid only when **Cluster Type** is set to **Custom**. For details, see **Custom Cluster Template Description**.
Node Type	MRS provides three types of nodes: ● Master: A Master node in an MRS cluster manages the cluster, assigns executable cluster files to Core nodes, traces the execution status of each job, and monitors the DataNode running status. ● Core: A Core node in a cluster processes data and stores process data in HDFS. Analysis Core nodes are created in an analysis cluster. Streaming Core nodes are created in a streaming cluster. Both analysis and streaming Core nodes are created in a hybrid cluster. ● Task: A Task node in a cluster is used for computing and does not store persistent data. Yarn and Storm are mainly installed on Task nodes. Task nodes are optional, and the number of Task nodes can be zero. Analysis Task nodes are created in an analysis cluster. Streaming Task nodes are created in a streaming cluster. Both analysis and streaming Task nodes are created in a hybrid cluster. When the data volume change is small in a cluster but the cluster's service processing capabilities need to be remarkably and temporarily improved, add Task nodes to address the following situations: – Service volumes temporarily increase, for example, report processing at the end of the year. – Long-term tasks must be completed in a short time, for example, some urgent analysis tasks.
Instance Specifications	Instance specifications of Master or Core nodes. MRS supports host specifications determined by CPU, memory, and disk space. Click ✎ to configure the instance specifications, system disk, and data disk parameters of the cluster node. **NOTE** ● More advanced instance specifications provide better data processing. ● For MRS 3.*x* or later, the memory of the master node must be greater than 64 GB.

Parameter	Description
System Disk	Storage type and storage space of the system disk on a node. Storage type can be any of the following: ● SATA: common I/O ● SAS: high I/O ● SSD: ultra-high I/O ● GPSSD: general-purpose SSD
Data Disk	Data disk storage space of a node. To increase data storage capacity, you can add disks at the same time when creating a cluster. The following two application scenarios are involved. ● Data storage and computing are separated. Data is stored in OBS, which features low cost and unlimited storage capacity. The clusters can be terminated at any time in OBS. The computing performance is determined by OBS access performance and is lower than that of HDFS. This configuration is recommended if data computing is infrequent. ● Data storage and computing are not separated. Data is stored in HDFS, which features high cost, high computing performance, and limited storage capacity. Before terminating clusters, you must export and store the data. This configuration is recommended if data computing is frequent. The storage type can be any of the following: ● SATA: common I/O ● SAS: high I/O ● SSD: ultra-high I/O ● GPSSD: general-purpose SSD **NOTE** More nodes in a cluster require higher disk capacity of Master nodes. To ensure stable cluster running, set the disk capacity of the Master node to over 600 GB if the number of nodes is 300 and increase it to over 1 TB if the number of nodes reaches 500.

Parameter	Description
Instance Count	Number of Master and Core nodes. For Master nodes: - If **Cluster HA** is enabled, the number of Master nodes is fixed to **2**. - If **Cluster HA** is disabled, the number of Master nodes is fixed to **1**. At least one Core node must exist and the total number of Core and Task nodes cannot exceed 500. Task: Click ⊕ to add a Task node. Click ✎ to modify the instance specifications and disk configuration of a Task node. Click 🗑 to delete the added Task node. **NOTE** - A maximum of 500 Core nodes are supported by default. If more than 500 Core nodes are required, contact technical support. - A small number of nodes may cause clusters to run slowly while a large number of nodes may be unnecessarily costly. Set an appropriate value based on data to be processed.
Topology Adjustment	If the deployment mode in the **Common Node** does not meet the requirements, set **Topology Adjustment** to **Enable** and adjust the instance deployment mode based on service requirements. For details, see **Topology Adjustment for a Custom Cluster**. This parameter is valid only when **Cluster Type** is set to **Custom**.

(Optional) Advanced Configuration

Table 3-4 MRS cluster advanced configuration topology

Parameter	Description
Tag	For details, see **Adding a Tag to a Cluster**.
Hostname Prefix	Enter the prefix for the computer hostname of an ECS in the cluster.
Auto Scaling	Auto scaling can be configured only after you specify Task node specifications in the **Configure Hardware** step. For details about how to configure Task node specifications, see **Configuring an Auto Scaling Rule**.
Bootstrap Action	For details, see **Adding a Bootstrap Action**. This parameter is not available in MRS 3.*x*.

Parameter	Description
Agency	By binding an agency, ECSs or BMSs can manage some of your resources. Determine whether to configure an agency based on the actual service scenario. For example, you can configure an agency of the ECS type to automatically obtain the AK/SK to access OBS. For details, see **Configuring a Storage-Compute Decoupled Cluster (Agency)**. The **MRS_ECS_DEFAULT_AGENCY** agency has the OBSOperateAccess permission of OBS and the CESFullAccess (for users who have enabled fine-grained policies), CES Administrator, and KMS Administrator permissions in the region where the cluster is located.
Metric Sharing	Monitoring metrics of big data components are collected. If a fault occurs when you use a cluster, share the monitoring metrics with technical support for troubleshooting. This parameter is not available in MRS 3.*x*.
OBS Permission Control	Users who have enabled fine-grained permission control can use this function to grant permissions on different directories in OBS file systems to different MRS users. For details, see **Configuring Fine-Grained Permissions for MRS Multi-User Access to OBS**. This parameter is not available in MRS 3.*x*.
Data Disk Encryption	Whether to encrypt data in the data disk mounted to the cluster. This function is disabled by default. To use this function, you must have the Security Administrator and KMS Administrator permissions. This parameter is not available in MRS 3.*x*. Keys used by encrypted data disks are provided by the Key Management Service (KMS) of the Data Encryption Workshop (DEW), secure and convenient. Therefore, you do not need to establish and maintain the key management infrastructure. Click **Data Disk Encryption** to enable or disable the data disk encryption function.
Key ID	This parameter is displayed only when the **Data Disk Encryption** function is enabled. This parameter indicates the key ID corresponding to the selected key name. This parameter is not available in MRS 3.*x*.

Parameter	Description
Key Name	This parameter is mandatory when the **Data Disk Encryption** function is enabled. Select the name of the key used to encrypt the data disk. By default, the default master key named **evs/default** is selected. You can select another master key from the drop-down list. This parameter is not available in MRS 3.*x*. If disks are encrypted using a CMK, which is then disabled or scheduled for deletion, the disks can no longer be read from or written to, and data on these disks may never be restored. Exercise caution when performing this operation. Click **View Key List** to enter a page where you can create and manage keys.
Alarm	If the alarm function is enabled, the cluster maintenance personnel can be notified in a timely manner to locate faults when the cluster runs abnormally or the system is faulty.
Rule Name	Name of the rule for sending alarm messages. The value can contain only digits, letters, hyphens (-), and underscores (_).
Topic Name	Select an existing topic or click **Create Topic** to create a topic. To deliver messages published to a topic, you need to add a subscriber to the topic. For details, see **Adding Subscriptions to a Topic**. A topic serves as a message sending channel, where publishers and subscribers can interact with each other.
Kerberos Authentication	Whether to enable Kerberos authentication when logging in to Manager. ● : If **Kerberos Authentication** is disabled, common users can use all functions of an MRS cluster. You are advised to disable Kerberos authentication in single-user scenarios. ● : If **Kerberos Authentication** is enabled, common users cannot use the file and job management functions of an MRS cluster and cannot view cluster resource usage or the job records for Hadoop and Spark. To use more cluster functions, the users must contact the Manager administrator to assign more permissions. You are advised to enable Kerberos authentication in multi-user scenarios.
Username	Name of the administrator of Manager. **admin** is used by default.

Parameter	Description
Password	Password of the Manager administrator The following requirements must be met: ● Must contain 8 to 26 characters. ● Must contain at least four of the following: – Lowercase letters – Uppercase letters – Digits – Have at least one of the following special characters: !?,.: -_{} [ ]@ $% ^ + = / ● Cannot be the same as the username or the username spelled backwards. Password Strength: The colorbar in red, orange, and green indicates weak, medium, and strong password, respectively.
Confirm Password	Enter the password of the Manager administrator again.
Key Pair	Key pairs are used to log in to ECS nodes of the cluster. Select a key pair from the drop-down list. Select "I acknowledge that I have obtained private key file *SSHkey-xxx* and that without this file I will not be able to log in to my ECS." If you have never created a key pair, click **View Key Pair** to create or import a key pair. And then, obtain a private key file. A key pair, also called an SSH key, consists of a public key and a private key. You can create an SSH key and download the private key for authenticating remote login. For security, a private key can only be downloaded once. Keep it secure. Use an SSH key in either of the following two methods: 1. Creating an SSH key: After you create an SSH key, a public key and a private key are generated. The public key is stored in the system, and the private key is stored in the local ECS. When you log in to an ECS, the public and private keys are used for authentication. 2. Importing an SSH key: If you have obtained the public and private keys, import the public key into the system. When you log in to an ECS, the public and private keys are used for authentication.

Parameter	Description
Secure Communications	MRS clusters provision, manage, and use big data components through the management console. Big data components are deployed in a user's VPC. If the MRS management console needs to directly access big data components deployed in the user's VPC, you need to enable the corresponding security group rules after you have obtained user authorization. This authorization process is called secure communications. For details, see **Communication Security Authorization**. If the secure communications function is not enabled, MRS clusters cannot be created.

Failed to Create a Cluster

If a cluster fails to be created, the failed task will be managed on the **Manage Failed Tasks** page. Choose **Clusters** > **Active Clusters**. Click 🔴 shown in **Figure 3-1** to go to the **Manage Failed Tasks** page. In the **Status** column, hover the cursor over 🔴 to view the failure cause. You can delete failed tasks by referring to **Viewing Failed MRS Tasks**.

Figure 3-1 Failed task management

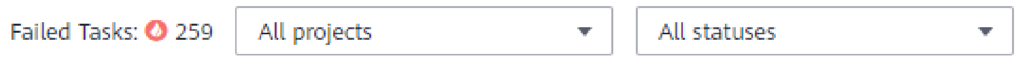


Table 3-5 lists the error codes of MRS cluster creation failures.

Table 3-5 Error codes

Error Code	Description
MRS.101	Insufficient quota to meet your request. Contact customer service to increase the quota.
MRS.102	The token cannot be null or invalid. Try again later or contact customer service.
MRS.103	Invalid request. Try again later or contact customer service.
MRS.104	Insufficient resources. Try again later or contact customer service.
MRS.105	Insufficient IP addresses in the existing subnet. Try again later or contact customer service.
MRS.201	Failed due to an ECS error. Try again later or contact customer service.

Error Code	Description
MRS.202	Failed due to an IAM error. Try again later or contact customer service.
MRS.203	Failed due to a VPC error. Try again later or contact customer service.
MRS.400	MRS system error. Try again later or contact customer service.

3.4 Creating a Custom Topology Cluster

The analysis cluster, streaming cluster, and hybrid cluster provided by MRS use fixed templates to deploy cluster processes. Therefore, you cannot customize service processes on management nodes and control nodes. If you want to customize the cluster deployment, set **Cluster Type** to **Custom** when creating a cluster. In this way, you can customize the deployment mode of process instances on the management nodes and control nodes in the cluster. Only MRS 3.x and later versions support the creation of clusters in a custom topology.

A custom cluster provides the following functions:

- Separated deployment of the management and control roles: The management role and control role are deployed on different Master nodes.
- Co-deployment of the management and control roles: The management and control roles are co-deployed on the Master node.
- ZooKeeper is deployed on an independent node to improve reliability.
- Components are deployed separately to avoid resource contention.

Roles in an MRS cluster:

- Management Node (MN): is the node to install Manager (the management system of the MRS cluster). It provides a unified access entry. Manager centrally manages nodes and services deployed in the cluster.
- Control Node (CN): controls and monitors how data nodes store and receive data, and send process status, and provides other public functions. Control nodes of MRS include HMaster, HiveServer, ResourceManager, NameNode, JournalNode, and SlapdServer.
- Data Node (DN): A data node executes the instructions sent by the management node, reports task status, stores data, and provides other public functions. Data nodes of MRS include DataNode, RegionServer, and NodeManager.

Customizing a Cluster

Step 1 Log in to the MRS console.

Step 2 Click **Create Cluster**. The page for creating a cluster is displayed.

Step 3 Click the **Custom Config** tab.

Step 4 Configure basic cluster information. For details about the parameters, see **Software Configurations**.

- **Region**: Retain the default value.
- **Cluster Name**: You can use the default name. However, you are advised to include a project name abbreviation or date for consolidated memory and easy distinguishing, for example, **mrs_20180321**.
- **Cluster Version**: Currently, only MRS 3.x are supported.
- **Cluster Type**: Select **Custom** and select components as required.

Step 5 Click **Next**. Configure hardware information.

- **AZ**: Retain the default value.
- **VPC**: Retain the default value. If there is no available VPC, click **View VPC** to access the VPC console and create a new VPC.
- **Subnet**: Retain the default value.
- **Security Group**: Select **Auto create**.
- **EIP**: Select **Bind later**.
- **Common Node**: For details, see **Custom Cluster Template Description**.
- **Instance Specifications**: Click ✎ to configure the instance specifications, system disk and data disk storage types, and storage space.
- **Instance Count**: Adjust the number of cluster instances based on the service volume. For details, see **Table 3-7**.
- **Topology Adjustment**: If the deployment mode in the **Common Node** does not meet the requirements, you need to manually install some instances that are not deployed by default, or you need to manually install some instances, set **Topology Adjustment** to **Enable** and adjust the instance deployment mode based on service requirements. For details, see **Topology Adjustment for a Custom Cluster**.

Step 6 Click **Next** and set advanced options.

For details about the parameters, see **(Optional) Advanced Configuration**.

Step 7 Click **Now**.

If Kerberos authentication is enabled for a cluster, check whether Kerberos authentication is required. If yes, click **Continue**. If no, click **Back** to disable Kerberos authentication and then create a cluster.

Step 8 Click **Back to Cluster List** to view the cluster status.

It takes some time to create a cluster. The initial status of the cluster is **Starting**. After the cluster has been created successfully, the cluster status becomes **Running**.

----End

Custom Cluster Template Description

Table 3-6 Common templates for custom clusters

Common Node	Description	Node Range
Compact	The management role and control role are deployed on the Master node, and data instances are deployed in the same node group. This deployment mode applies to scenarios where the number of control nodes is less than 100, reducing costs.	● The number of Master nodes is greater than or equal to 3 and less than or equal to 11. ● The total number of node groups is less than or equal to 10, and the total number of nodes in non-Master node groups is less than or equal to 10,000.
OMS-separate	The management role and control role are deployed on different Master nodes, and data instances are deployed in the same node group. This deployment mode is applicable to a cluster with 100 to 500 nodes and delivers better performance in high-concurrency load scenarios.	● The number of Master nodes is greater than or equal to 5 and less than or equal to 11. ● The total number of node groups is less than or equal to 10, and the total number of nodes in non-Master node groups is less than or equal to 10,000.
Full-size	The management role and control role are deployed on different Master nodes, and data instances are deployed in different node groups. This deployment mode is applicable to a cluster with more than 500 nodes. Components can be deployed separately, which can be used for a larger cluster scale.	● The number of Master nodes is greater than or equal to 9 and less than or equal to 11. ● The total number of node groups is less than or equal to 10, and the total number of nodes in non-Master node groups is less than or equal to 10,000.

Table 3-7 Node deployment scheme of a customized MRS cluster

Node Deployment Principle		Applicable Scenario	Networking Rule
Management nodes, control nodes,	MN × 2 + CN × 9 + DN × n	(Recommended) This scheme is used when the number of data nodes is 500–2000.	● If the number of nodes in a cluster exceeds 200, the nodes are distributed to different subnets and

Node Deployment Principle		Applicable Scenario	Networking Rule
and data nodes are deployed separately. (This scheme requires at least eight nodes.)	MN × 2 + CN × 5 + DN × n	(Recommended) This scheme is used when the number of data nodes is 100–500.	the subnets are interconnected with each other in Layer 3 using core switches. Each subnet can contain a maximum of 200 nodes and the allocation of nodes to different subnets must be balanced. ● If the number of nodes is less than 200, the nodes in the cluster are deployed in the same subnet and the nodes are interconnected with each other in Layer 2 using aggregation switches.
	MN × 2 + CN × 3 + DN × n	(Recommended) This scheme is used when the number of data nodes is 30–100.	
The management nodes and control nodes are deployed together, and the data nodes are deployed separately.	(MN+CN) × 3 + DN × n	(Recommended) This scheme is used when the number of data nodes is 3–30.	Nodes in the cluster are deployed in the same subnet and are interconnected with each other at Layer 2 through aggregation switches.

Node Deployment Principle	Applicable Scenario	Networking Rule
The management nodes, control nodes, and data nodes are deployed together.	● This scheme is applicable to a cluster having fewer than 6 nodes. ● This scheme requires at least three nodes. **NOTE** This template is not recommended in the production environment or commercial environment. ● If management, control, and data nodes are co-deployed, cluster performance and reliability are greatly affected. ● If the number of nodes meet the requirements, deploy data nodes separately. ● If the number of nodes is insufficient to support separately deployed data nodes, use the dual-plane networking mode for this scenario. The traffic of the management network is isolated from that of the service network to prevent excessive data volumes on the service plane, ensuring correct delivery of management operations.	Nodes in the cluster are deployed in the same subnet and are interconnected with each other at Layer 2 through aggregation switches.

Topology Adjustment for a Custom Cluster

Table 3-8 Topology adjustment

Service	Dependency	Role	Role Deployment Suggestions	Description
OMSServer	-	OMSServer	This role can be deployed it on the Master node and cannot be modified.	-

Service	Dependency	Role	Role Deployment Suggestions	Description
ClickHouse	Depends on ZooKeeper.	CHS (ClickHouseServer)	This role can be deployed on all nodes. Number of role instances to be deployed: an even number ranging from 2 to 256	A non-Master node group with this role assigned is considered as a Core node.
		CLB (ClickHouseBalancer)	This role can be deployed on all nodes. Number of role instances to be deployed: 2 to 256	-
ZooKeeper	-	QP(quorumpeer)	This role can be deployed on the Master node only. Number of role instances to be deployed: 3 to 9, with the step size of 2	-
Hadoop	Depends on ZooKeeper.	NN(NameNode)	This role can be deployed on the Master node only. Number of role instances to be deployed: 2	The NameNode and ZKFC processes are deployed on the same server for cluster HA.
		HFS (HttpFS)	This role can be deployed on the Master node only. Number of role instances to be deployed: 0 to 10	-
		JN(JournalNode)	This role can be deployed on the Master node only. Number of role instances to be deployed: 3 to 60, with the step size of 2	-

Service	Dependency	Role	Role Deployment Suggestions	Description
		DN(Data Node)	This role can be deployed on all nodes. Number of role instances to be deployed: 3 to 10,000	A non-Master node group with this role assigned is considered as a Core node.
		RM(ResourceManager)	This role can be deployed on the Master node only. Number of role instances to be deployed: 2	-
		NM(NodeManager)	This role can be deployed on all nodes. Number of role instances to be deployed: 3 to 10,000	-
		JHS(JobHistoryServer)	This role can be deployed on the Master node only. Number of role instances to be deployed: 1 to 2	-
		TLS(TimelineServer)	This role can be deployed on the Master node only. Number of role instances to be deployed: 0 to 1	-
Presto	Depends on Hive.	PCD(Coordinator)	This role can be deployed on the Master node only. Number of role instances to be deployed: 2	-

Service	Dependency	Role	Role Deployment Suggestions	Description
		PWK(Worker)	This role can be deployed on all nodes. Number of role instances to be deployed: 1 to 10,000	-
Spark2x	● Depends on Hadoop. ● Depends on Hive. ● Depends on ZooKeeper.	JS2X(JDBCServer2x)	This role can be deployed on the Master node only. Number of role instances to be deployed: 2 to 10	-
		JH2X(JobHistory2x)	This role can be deployed on the Master node only. Number of role instances to be deployed: 2	-
		SR2X(SparkResource2x)	This role can be deployed on the Master node only. Number of role instances to be deployed: 2 to 50	-
		IS2X(IndexServer2x)	(Optional) This role can be deployed on the Master node only. Number of role instances to be deployed: 0 to 2, with the step size of 2	-
HBase	Depends on Hadoop.	HM(HMaster)	This role can be deployed on the Master node only. Number of role instances to be deployed: 2	-

Service	Dependency	Role	Role Deployment Suggestions	Description
		TS(ThriftServer)	This role can be deployed on all nodes. Number of role instances to be deployed: 0 to 10,000	-
		RT(RESTServer)	This role can be deployed on all nodes. Number of role instances to be deployed: 0 to 10,000	-
		RS(RegionServer)	This role can be deployed on all nodes. Number of role instances to be deployed: 3 to 10,000	-
		TS1(Thrift1Server)	This role can be deployed on all nodes. Number of role instances to be deployed: 0 to 10,000	If the Hue service is installed in a cluster and HBase needs to be used on the Hue web UI, install this instance for the HBase service.
Hive	● Depends on Hadoop. ● Depends on DBService.	MS(MetaStore)	This role can be deployed on the Master node only. Number of role instances to be deployed: 2 to 10	-
		WH (WebHCat)	This role can be deployed on the Master node only. Number of role instances to be deployed: 1 to 10	-

Service	Dependency	Role	Role Deployment Suggestions	Description
		HS(HiveServer)	This role can be deployed on the Master node only. Number of role instances to be deployed: 2 to 80	-
Hue	Depends on DBService.	H(Hue)	This role can be deployed on the Master node only. Number of role instances to be deployed: 2	-
Sqoop	Depends on Hadoop.	SC(Sqoop Client)	This role can be deployed on all nodes. Number of role instances to be deployed: 1 to 10,000	-
Kafka	Depends on ZooKeeper.	B(Broker)	This role can be deployed on all nodes. Number of role instances to be deployed: 3 to 10,000	-
Flume	-	MS(MonitorServer)	This role can be deployed on the Master node only. Number of role instances to be deployed: 1 to 2	-
		F(Flume)	This role can be deployed on all nodes. Number of role instances to be deployed: 1 to 10,000	A non-Master node group with this role assigned is considered as a Core node.

Service	Dependency	Role	Role Deployment Suggestions	Description
Tez	- Depends on Hadoop. - Depends on DBService. - Depends on ZooKeeper.	TUI(TezUI)	This role can be deployed on the Master node only. Number of role instances to be deployed: 1 to 2	-
Flink	- Depends on ZooKeeper. - Depends on Hadoop.	FR(FlinkResource)	This role can be deployed on all nodes. Number of role instances to be deployed: 1 to 10,000	-
		FS(FlinkServer)	This role can be deployed on all nodes. Number of role instances to be deployed: 0 to 2	-
Oozie	- Depends on Hadoop. - Depends on DBService. - Depends on ZooKeeper.	O(oozie)	This role can be deployed on the Master node only. Number of role instances to be deployed: 2	-
Impala	- Depends on Hadoop. - Depends on Hive.	StateStore	This role can be deployed on the Master node only. Number of role instances to be deployed: 1	-

Service	Dependency	Role	Role Deployment Suggestions	Description
	- Depends on DBService. - Depends on ZooKeeper.	Catalog	This role can be deployed on the Master node only. Number of role instances to be deployed: 1	-
		Impalad	This role can be deployed on all nodes. Number of role instances to be deployed: 1 to 10,000	-
Kudu	-	KuduMaster	This role can be deployed on the Master node only. Number of role instances to be deployed: 3 or 5	-
		KuduTserver	This role can be deployed on all nodes. Number of role instances to be deployed: 3 to 10,000	-
Ranger	Depends on DBService.	RA(RangerAdmin)	This role can be deployed on the Master node only. Number of role instances to be deployed: 1 to 2	-
		USC(UserSync)	This role can be deployed on the Master node only. Number of role instances to be deployed: 1	-

Service	Dependency	Role	Role Deployment Suggestions	Description
		TSC (TagSync)	This role can be deployed on all nodes. Number of role instances to be deployed: 0 to 1	-

3.5 Adding a Tag to a Cluster

Tags are used to identify clusters. Adding tags to clusters can help you identify and manage your cluster resources.

You can add a maximum of 10 tags to a cluster when creating the cluster or add them on the details page of the created cluster.

A tag consists of a tag key and a tag value. **Table 3-9** provides tag key and value requirements.

Table 3-9 Tag key and value requirements

Parameter	Requirement	Example
Key	A tag key cannot be left blank. A tag key must be unique in a cluster. A tag key contains a maximum of 36 characters. A tag value cannot contain special characters (=*<>\,|/) or start or end with spaces.	Organization
Value	A tag value contains a maximum of 43 characters. A tag value cannot contain special characters (=*<>\,|/) or start or end with spaces. This parameter can be left blank.	Apache

Adding Tags to a Cluster

You can perform the following operations to add tags to a cluster when creating the cluster.

1. Log in to the MRS console.
2. Click **Create Cluster**. The corresponding page is displayed.
3. Click the **Custom Config** tab.
4. Configure the cluster software and hardware by referring to **Creating a Custom Cluster**.
5. On the **Set Advanced Options** tab page, add a tag.

   Enter the key and value of a tag to be added.

   You can add a maximum of 10 tags to a cluster and use intersections of tags to search for the target cluster.

   NOTE

   You can also add tags to existing clusters. For details, see **Managing Tags**.

Searching for the Target Cluster

On the **Active Clusters** page, search for the target cluster by tag key or tag value.

1. Log in to the MRS console.
2. In the upper right corner of the **Active Clusters** page, click **Search by Tag** to access the search page.
3. Enter the tag of the cluster to be searched.

   You can select a tag key or tag value from their drop-down lists. When the tag key or tag value is exactly matched, the system can automatically locate the target cluster. If you enter multiple tags, their intersections are used to search for the cluster.
4. Click **Search**.

   The system searches for the target cluster by tag key or value.

Managing Tags

You can view, add, modify, and delete tags on the **Tags** tab page of the cluster.

1. Log in to the MRS console.
2. On the **Active Clusters** page, click the name of a cluster for which you want to manage tags.

   The cluster details page is displayed.
3. Click the **Tags** tab and view, add, modify, and delete tags on the tab page.
   - View

     On the **Tags** tab page, you can view details about tags of the cluster, including the number of tags and the key and value of each tag.
   - Add

     Click **Add Tag** in the upper left corner. In the displayed **Add Tag** dialog box, enter the key and value of the tag to be added, and click **OK**.

- Modify

  In the **Operation** column of the tag, click **Edit**. In the displayed **Edit Tag** page, enter new tag key and value and click **OK**.

- Delete

  In the **Operation** column of the tag, click **Delete**. After confirmation, click **OK** in the displayed page for deleting a tag.

3.6 Communication Security Authorization

MRS clusters provision, manage, and use big data components through the management console. Big data components are deployed in a user's VPC. If the MRS management console needs to directly access big data components deployed in the user's VPC, you need to enable the corresponding security group rules after you have obtained user authorization. This authorization process is called secure communications.

If the secure communications function is not enabled, MRS clusters cannot be created. If you disable the communication after a cluster is created, the cluster status will be **Network channel is not authorized** and the following functions will be affected:

- Functions, such as big data component installation, cluster scale-out/scale-in, and Master node specification upgrade, are unavailable.
- The cluster running status, alarms, and events cannot be monitored.
- The node management, component management, alarm management, file management, job management, patch management, and tenant management functions on the cluster details page are unavailable.
- The Manager page and the website of each component cannot be accessed.

After the secure communications function is enabled again, the cluster status is restored to **Running**, and the preceding functions become available. For details, see **Enabling Secure Communications for Clusters with This Function Disabled**.

If the security group rules authorized in the cluster are insufficient for you to provision, manage, and use big data components, ❗ is displayed on the right of **Secure Communications**. In this case, click **Update** to update the security group rules. For details, see **Update**.

Enabling Secure Communications During Cluster Creation

Step 1 Log in to the MRS console.

Step 2 Click Create **Cluster**. The corresponding page is displayed.

Step 3 Click **Quick Config** or **Custom Config**.

Step 4 Configure cluster information by referring to **Creating a Custom Cluster**.

Step 5 In the **Secure Communications** area of the **Advanced Settings** tab page, select **Enable**.

Step 6 Click **Create Now**.

If Kerberos authentication is enabled for a cluster, check whether Kerberos authentication is required. If yes, click **Continue**. If no, click **Back** to disable Kerberos authentication and then create a cluster.

----End

Disabling Secure Communications After a Cluster Is Created

Step 1 Log in to the MRS console.

Step 2 In the active cluster list, click the name of the cluster for which you want to disable secure communications.

The cluster details page is displayed.

Step 3 Click the switch on the right of **Secure Communications** to disable authorization. In the dialog box that is displayed, click **OK**.

After the authorization is disabled, the cluster status changes to **Network channel unauthorized**, and some functions of the cluster are unavailable. Exercise caution when performing this operation.

----End

Enabling Secure Communications for Clusters with This Function Disabled

Step 1 Log in to the MRS console.

Step 2 In the active cluster list, click the name of the cluster for which you want to enable secure communications.

The cluster details page is displayed.

Step 3 Click the switch on the right of **Secure Communications** to enable the function.

After the function is enabled, the cluster status changes to **Running**.

----End

Update

If the security group rules authorized in the cluster are insufficient for you to provision, manage, and use big data components, ❗ is displayed on the right of **Secure Communications**. In this case, click **Update** to update the security group rules. For details, see **Update**.

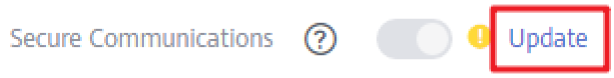
Step 1 Log in to the MRS console.

Step 2 In the active cluster list, click the name of the cluster for which you want to update secure communications.

The cluster details page is displayed.

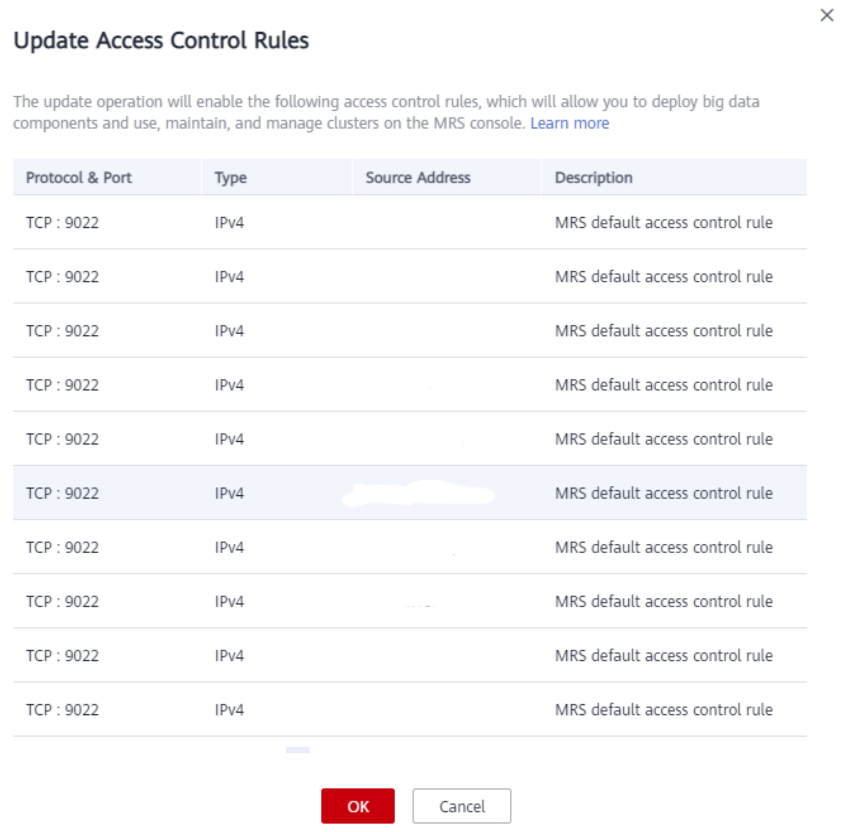
Step 3 Click **Update** on the right of **Secure Communications**.

Figure 3-2 Update



Step 4 Click **OK**.

Figure 3-3 Updating access control rules



Update Access Control Rules

The update operation will enable the following access control rules, which will allow you to deploy big data components and use, maintain, and manage clusters on the MRS console. Learn more

Protocol & Port	Type	Source Address	Description
TCP : 9022	IPv4		MRS default access control rule
TCP : 9022	IPv4		MRS default access control rule
TCP : 9022	IPv4		MRS default access control rule
TCP : 9022	IPv4		MRS default access control rule
TCP : 9022	IPv4		MRS default access control rule
TCP : 9022	IPv4		MRS default access control rule
TCP : 9022	IPv4		MRS default access control rule
TCP : 9022	IPv4		MRS default access control rule
TCP : 9022	IPv4		MRS default access control rule
TCP : 9022	IPv4		MRS default access control rule

----End

3.7 Configuring an Auto Scaling Rule

Background

In big data application scenarios, especially real-time data analysis and processing, the number of cluster nodes needs to be dynamically adjusted according to data volume changes to provide the required number of resources. The auto scaling function of MRS enables the task nodes of a cluster to be automatically scaled to match cluster loads. If the data volume changes periodically, you can configure an auto scaling rule so that the number of task nodes can be automatically adjusted in a fixed period of time before the data volume changes.

- Auto scaling rules: You can increase or decrease task nodes based on real-time cluster loads. Auto scaling will be triggered with a certain delay when the data volume changes.
- Resource plans: Set the task node quantity based on the time range. If the data volume changes periodically, you can create resource plans to resize the cluster before the data volume changes, thereby avoiding delays in increasing or decreasing resources.

You can configure either auto scaling rules or resource plans or both to trigger auto scaling. Configuring both resource plans and auto scaling rules improves the cluster node scalability to cope with occasionally unexpected data volume peaks.

In some service scenarios, resources need to be reallocated or service logic needs to be modified after cluster scale-out or scale-in. If you manually scale out or scale in a cluster, you can log in to cluster nodes to reallocate resources or modify service logic. If you use auto scaling, MRS enables you to customize automation scripts for resource reallocation and service logic modification. Automation scripts can be executed before and after auto scaling and automatically adapt to service load changes, all of which eliminates manual operations. In addition, automation scripts can be fully customized and executed at various moments, meeting your personalized requirements and improving auto scaling flexibility.

- Auto scaling rules:
  - You can set a maximum of five rules for scaling out or in a cluster, respectively.
  - The system determines the scale-out and then scale-in based on your configuration sequence. Important policies take precedence over other policies to prevent repeated triggering when the expected effect cannot be achieved after a scale-out or scale-in.
  - Comparison factors include greater than, greater than or equal to, less than, and less than or equal to.
  - Cluster scale-out or scale-in can be triggered only after the configured metric threshold is reached for consecutive 5*n* (the default value of *n* is 1) minutes.
  - After each scale-out or scale-in, there is a cooling duration is greater than 0, and lasts 20 minutes by defaults.
  - In each cluster scale-out or scale-in, at least one node and at most 100 nodes can be added or reduced.
- Resource plans (setting the number of Task nodes by time range):
  - You can specify a Task node range (minimum number to maximum number) in a time range. If the number of Task nodes is beyond the Task node range in a resource plan, the system triggers cluster scale-out or scale-in.
  - You can set a maximum of five resource plans for a cluster.
  - A resource plan cycle is by day. The start time and end time can be set to any time point between 00:00 and 23:59. The start time must be at least 30 minutes earlier than the end time. Time ranges configured for different resource plans cannot overlap.
  - After a resource plan triggers cluster scale-out or scale-in, there is 10-minute cooling duration. Auto scaling will not be triggered again within the cooling time.
  - When a resource plan is enabled, the number of Task nodes in the cluster is limited to the default node range configured by you in other time periods except the time period configured in the resource plan.
  - If the resource plan is not enabled, the number of Task nodes is not limited to the default node range.
- Automation scripts:

- You can set an automation script so that it can automatically run on cluster nodes when auto scaling is triggered.
- You can set a maximum number of 10 automation scripts for a cluster.
- You can specify an automation script to be executed on one or more types of nodes.
- Automation scripts can be executed before or after scale-out or scale-in.
- Before using automation scripts, upload them to a cluster VM or OBS file system in the same region as the cluster. The automation scripts uploaded to the cluster VM can be executed only on the existing nodes. If you want to make the automation scripts run on the new nodes, upload them to the OBS file system.

Accessing the Auto Scaling Configuration Page

You can configure an auto scaling rule on the **Set Advanced Options** page during cluster creation or on the **Nodes** page after the cluster is created.

Configuring an auto scaling rule when creating a cluster

Step 1 Log in to the MRS console.

Step 2 When you create a cluster containing task nodes, configure the cluster software and hardware information by referring to **Creating a Custom Cluster**. Then, on the **Set Advanced Options** page, enable **Analysis Task** and configure or modify auto scaling rules and resource plans.

You can configure the auto scaling rules by referring to the following scenarios:

- **Scenario 1: Using Auto Scaling Rules Alone**
- **Scenario 2: Using Resource Plans Alone**
- **Scenario 3: Using Both Auto Scaling Rules and Resource Plans**

----End

Configure an auto scaling rule for an existing cluster

Step 1 Log in to the MRS console.

Step 2 In the navigation pane on the left, choose **Clusters** > **Active Clusters** and click the name of a running cluster to go to the cluster details page.

Step 3 Click the **Nodes** tab and then **Auto Scaling** in the **Operation** column of the task node group. The **Auto Scaling** page is displayed.

NOTE

- If no task node exists in the cluster, click **Configure Task Node** to add one and then configure the auto scaling rules.
- For MRS 3.*x* or later, **Configure Task Node** is available only for analysis clusters, streaming clusters, and hybrid clusters. For details about how to add a task node for a custom cluster of MRS 3.*x* or later, see **Adding a Task Node**.

Step 4 Enable **Auto Scaling** and configure or modify auto scaling rules and resource plans.

You can configure the auto scaling rules by referring to the following scenarios:

- **Scenario 1: Using Auto Scaling Rules Alone**
- **Scenario 2: Using Resource Plans Alone**
- **Scenario 3: Using Both Auto Scaling Rules and Resource Plans**

----End

Scenario 1: Using Auto Scaling Rules Alone

The following is an example scenario:

The number of nodes needs to be dynamically adjusted based on the Yarn resource usage. When the memory available for Yarn is less than 20% of the total memory, five nodes need to be added. When the memory available for Yarn is greater than 70% of the total memory, five nodes need to be removed. The number of nodes in a task node group ranges from 1 to 10.

Step 1 Go to the **Auto Scaling** page to configure auto scaling rules.

- Configure the **Default Range** parameter.

  Enter a task node range, in which auto scaling is performed. This constraint applies to all scale-in and scale-out rules. The maximum value range allowed is 0 to 500.

  The value range in this example is 1 to 10.

- Configure an auto scaling rule.

  To enable **Auto Scaling**, you must configure a scale-out or scale-in rule.

  a. Select **Scale-Out** or **Scale-In**.

  b. Click **Add Rule**.

  c. Configure the **Rule Name**, **If**, **Last for**, **Add**, and **Cooldown Period** parameters.

  d. Click **OK**.

     You can view, edit, or delete the rules you configured in the **Scale-out** or **Scale-in** area on the **Auto Scaling** page. You can click **Add Rule** to configure multiple rules.

Step 2 (Optional) Configure automation scripts.

Set **Advanced Settings** to **Configure** and click **Created**, or click **Add Automation Script** to go to the **Automation Script** page.

MRS 3.*x* does not support this operation.

1. Set the following parameters: **Name**, **Script Path**, **Execution Node**, **Parameter**, **Executed**, and **Action upon Failure**. For details about the parameters, see **Table 3-12**.
2. Click **OK** to save the automation script configurations.

Step 3 Click **OK**.

NOTE

If you want to configure an auto scaling rule for an existing cluster, select **I agree to authorize MRS to scale out or in nodes based on the above rule**.

----End

Scenario 2: Using Resource Plans Alone

If the data volume changes regularly every day and you want to scale out or in a cluster before the data volume changes, you can create resource plans to adjust the number of Task nodes as planned in the specified time range.

Example:

A real-time processing service sees a sharp increase in data volume from 7:00 to 13:00 every day. Assume that an MRS streaming cluster is used to process the service data. Five task nodes are required from 7:00 to 13:00, while only two are required at other time.

Step 1 Go to the **Auto Scaling** page to configure a resource plan.

1. For example, the **Default Range** is set to **2-2**, indicating that the number of Task nodes is fixed to 2 except the time range specified in the resource plan.
2. Click **Configure Node Range for Specific Time Range** under **Default Range** or **Add Resource Plan**.
3. Configure **Time Range** and **Node Range**.

   For example, set **Time Range** to **07:00-13:00**, and **Node Range** to **5-5**. This indicates that the number of task nodes is fixed at 5 from 07:00 to 13:00.

   For details about parameter configurations, see **Table 3-11**. You can click **Configure Node Range for Specific Time Range** to configure multiple resource plans.

   NOTE

   – If you do not set **Node Range**, its default value will be used.
   – If you set both **Node Range** and **Time Rang**e, the node range you set will be used during the time range you set, and the default node range will be used beyond the time range you set. If the time is not within the configured time range, the default range is used.

Step 2 (Optional) Configure automation scripts.

Set **Advanced Settings** to **Configure** and click **Created**, or click **Add Automation Script** to go to the **Automation Script** page.

MRS 3.*x* does not support this operation.

1. Set the following parameters: **Name**, **Script Path**, **Execution Node**, **Parameter**, **Executed**, and **Action upon Failure**. For details about the parameters, see **Table 3-12**.
2. Click **OK** to save the automation script configurations.

Step 3 Click **OK**.

NOTE

If you want to configure an auto scaling rule for an existing cluster, select **I agree to authorize MRS to scale out or in nodes based on the above rule**.

----End

Scenario 3: Using Both Auto Scaling Rules and Resource Plans

If the data volume is not stable and the expected fluctuation may occur, the fixed Task node range cannot guarantee that the requirements in some service scenarios are met. In this case, it is necessary to adjust the number of Task nodes based on the real-time loads and resource plans.

The following is an example scenario:

A real-time processing service sees an unstable increase in data volume from 7:00 to 13:00 every day. For example, 5 to 8 task nodes are required from 7:00 to 13:00, and 2 to 4 are required beyond this period. Therefore, you can set an auto scaling rule based on a resource plan. When the data volume exceeds the expected value, the number of Task nodes can be adjusted if resource loads change, without exceeding the node range specified in the resource plan. When a resource plan is triggered, the number of nodes is adjusted within the specified node range with minimum affect. That is, increase nodes to the upper limit and decrease nodes to the lower limit.

Step 1 Go to the **Auto Scaling** page to configure auto scaling rules.

- **Default Range**

  Enter a task node range, in which auto scaling is performed. This constraint applies to all scale-in and scale-out rules.

  For example, this parameter is set to **2-4** in this scenario.

- **Auto Scaling**

  To enable **Auto Scaling**, you must configure a scale-out or scale-in rule.

  a. Select **Scale-Out** or **Scale-In**.

  b. Click **Add Rule**. The **Add Rule** page is displayed.

  c. Configure the **Rule Name**, **If**, **Last for**, **Add**, and **Cooldown Period** parameters.

  d. Click **OK**.

     You can view, edit, or delete the rules you configured in the **Scale-out** or **Scale-in** area on the **Auto Scaling** page.

Step 2 Configure a resource plan.

1. Click **Configure Node Range for Specific Time Range** under **Default Range** or **Add Resource Plan**.
2. Configure **Time Range** and **Node Range**.

   For example, **Time Range** is set to **07:00-13:00** and **Node Range** to **5-8**.

   For details about parameter configurations, see Table 3-11. You can click **Configure Node Range for Specific Time Range** or **Add Resource Plan** to configure multiple resource plans.

   NOTE

   - If you do not set **Node Range**, its default value will be used.
   - If you set both **Node Range** and **Time Rang**e, the node range you set will be used during the time range you set, and the default node range will be used beyond the time range you set. If the time is not within the configured time range, the default range is used.

Step 3 (Optional) Configure automation scripts.

Set **Advanced Settings** to **Configure** and click **Created**, or click **Add Automation Script** to go to the **Automation Script** page.

MRS 3.*x* does not support this operation.

1. Set the following parameters: **Name**, **Script Path**, **Execution Node**, **Parameter**, **Executed**, and **Action upon Failure**. For details about the parameters, see **Table 3-12**.
2. Click **OK** to save the automation script configurations.

Step 4 Click **OK**.

NOTE

If you want to configure an auto scaling rule for an existing cluster, select **I agree to authorize MRS to scale out or in nodes based on the above rule**.

----End

Related Information

When adding a rule, you can refer to **Table 3-10** to configure the corresponding metrics.

Table 3-10 Auto scaling metrics

Cluster Type	Metric	Value Type	Description
Streaming cluster	StormSlotAvailable	Integer	Number of available Storm slots Value range: 0 to 2147483646
	StormSlotAvailablePercentage	Percentage	Percentage of available Storm slots, that is, the proportion of the available slots to total slots Value range: 0 to 100
	StormSlotUsed	Integer	Number of the used Storm slots Value range: 0 to 2147483646
	StormSlotUsedPercentage	Percentage	Percentage of the used Storm slots, that is, the proportion of the used slots to total slots Value range: 0 to 100
	StormSupervisorMemAverageUsage	Integer	Average memory usage of the Supervisor process of Storm Value range: 0 to 2147483646
	StormSupervisorMemAverageUsagePercentage	Percentage	Average percentage of the used memory of the Supervisor process of Storm to the total memory of the system Value range: 0 to 100

Cluster Type	Metric	Value Type	Description
	StormSupervisorCPUAverageUsagePercentage	Percentage	Average percentage of the used CPUs of the Supervisor process of Storm to the total CPUs Value range: 0 to 6000
Analysis cluster	YARNAppPending	Integer	Number of pending tasks on YARN Value range: 0 to 2147483646
	YARNAppPendingRatio	Ratio	Ratio of pending tasks on Yarn, that is, the ratio of pending tasks to running tasks on Yarn Value range: 0 to 2147483646
	YARNAppRunning	Integer	Number of running tasks on Yarn Value range: 0 to 2147483646
	YARNContainerAllocated	Integer	Number of containers allocated to Yarn Value range: 0 to 2147483646
	YARNContainerPending	Integer	Number of pending containers on Yarn Value range: 0 to 2147483646
	YARNContainerPendingRatio	Ratio	Ratio of pending containers on Yarn, that is, the ratio of pending containers to running containers on Yarn. Value range: 0 to 2147483646
	YARNCPUAllocated	Integer	Number of virtual CPUs (vCPUs) allocated to Yarn Value range: 0 to 2147483646
	YARNCPUAvailable	Integer	Number of available vCPUs on Yarn Value range: 0 to 2147483646
	YARNCPUAvailablePercentage	Percentage	Percentage of available vCPUs on Yarn, that is, the proportion of available vCPUs to total vCPUs Value range: 0 to 100
	YARNCPUPending	Integer	Number of pending vCPUs on Yarn Value range: 0 to 2147483646

Cluster Type	Metric	Value Type	Description
	YARNMemoryAllocated	Integer	Memory allocated to Yarn. The unit is MB. Value range: 0 to 2147483646
	YARNMemoryAvailable	Integer	Available memory on Yarn. The unit is MB. Value range: 0 to 2147483646
	YARNMemoryAvailablePercentage	Percentage	Percentage of available memory on Yarn, that is, the proportion of available memory to total memory on Yarn Value range: 0 to 100
	YARNMemoryPending	Integer	Pending memory on Yarn Value range: 0 to 2147483646

NOTE

- When the value type is percentage or ratio in **Table 3-10**, the valid value can be accurate to percentile. The percentage metric value is a decimal value with a percent sign (%) removed. For example, 16.80 represents 16.80%.
- Hybrid clusters support all metrics of analysis and streaming clusters.

When adding a resource plan, you can set parameters by referring to **Table 3-11**.

Table 3-11 Configuration items of a resource plan

Configuration Item	Description
Time Range	Start time and End time of a resource plan are accurate to minutes, with the value ranging from **00:00** to **23:59**. For example, if a resource plan starts at 8:00 and ends at 10:00, set this parameter to 8:00-10:00. The end time must be at least 30 minutes later than the start time.
Node Range	The number of nodes in a resource plan ranges from **0** to **500**. In the time range specified in the resource plan, if the number of Task nodes is less than the specified minimum number of nodes, it will be increased to the specified minimum value of the node range at a time. If the number of Task nodes is greater than the maximum number of nodes specified in the resource plan, the auto scaling function reduces the number of Task nodes to the maximum value of the node range at a time. The minimum number of nodes must be less than or equal to the maximum number of nodes.

NOTE

- When a resource plan is enabled, the **Default Range** value on the auto scaling page forcibly takes effect beyond the time range specified in the resource plan. For example, if **Default Range** is set to **1-2**, **Time Range** is between **08:00-10:00**, and **Node Range** is **4-5** in a resource plan, the number of Task nodes in other periods (0:00-8:00 and 10:00-23:59) of a day is forcibly limited to the default node range (1 to 2). If the number of nodes is greater than 2, auto scale-in is triggered; if the number of nodes is less than 1, auto scale-out is triggered.
- When a resource plan is not enabled, the **Default Range** takes effect in all time ranges. If the number of nodes is not within the default node range, the number of Task nodes is automatically increased or decreased to the default node range.
- Time ranges of resource plans cannot be overlapped. The overlapped time range indicates that two effective resource plans exist at a time point. For example, if resource plan 1 takes effect from **08:00** to **10:00** and resource plan 2 takes effect from **09:00** to **11:00**, the time range between **09:00** to **10:00** is overlapped.
- The time range of a resource plan must be on the same day. For example, if you want to configure a resource plan from **23:00** to **01:00** in the next day, configure two resource plans whose time ranges are **23:00-00:00** and **00:00-01:00**, respectively.

When adding an automation script, you can set related parameters by referring to **Table 3-12**.

Table 3-12 Configuration items of an automation script

Configuration Item	Description
Name	Automation script name. The value can contain only digits, letters, spaces, hyphens (-), and underscores (_) and must not start with a space. The value can contain 1 to 64 characters. **NOTE** A name must be unique in the same cluster. You can set the same name for different clusters.
Script Path	Script path. The value can be an OBS file system path or a local VM path. ● An OBS file system path must start with **s3a://** and end with **.sh**, for example, **s3a://mrs-samples/*xxx*.sh**. ● A local VM path must start with a slash (/) and end with **.sh**. For example, the path of the example script for installing the Zepelin is **/opt/bootstrap/zepelin/zepelin_install.sh**.
Execution Node	Select a type of the node where an automation script is executed. **NOTE** ● If you select **Master** nodes, you can choose whether to run the script only on the active Master nodes by enabling or disabling the **Active Master** switch. ● If you enable it, the script runs only on the active Master nodes. If you disable it, the script runs on all Master nodes. This switch is disabled by default.

Configuration Item	Description
Parameter	Automation script parameter. The following predefined variables can be imported to obtain auto scaling information: ● **${mrs_scale_node_num}**: Number of auto scaling nodes. The value is always positive. ● **${mrs_scale_type}**: Scale-out/in type. The value can be **scale_out** or **scale_in**. ● **${mrs_scale_node_hostnames}**: Host names of the auto scaling nodes. Use commas (,) to separate multiple host names. ● **${mrs_scale_node_ips}**: IP address of the auto scaling nodes. Use commas (,) to separate multiple IP addresses. ● **${mrs_scale_rule_name}**: Name of the triggered auto scaling rule. For a resource plan, this parameter is set to **resource_plan**.
Executed	Time for executing an automation script. The following four options are supported: **Before scale-out**, **After scale-out**, **Before scale-in**, and **After scale-in**. **NOTE** Assume that the execution nodes include Task nodes. ● The automation script executed before scale-out cannot run on the Task nodes to be added. ● The automation script executed after scale-out can run on the added Task nodes. ● The automation script executed before scale-in can run on Task nodes to be deleted. ● The automation script executed after scale-in cannot run on the deleted Task nodes.
Action upon Failure	Whether to continue to execute subsequent scripts and scale-out/in after the script fails to be executed. **NOTE** ● You are advised to set this parameter to **Continue** in the commissioning phase so that the cluster can continue the scale-out/in operation no matter whether the script is executed successfully. ● If the script fails to be executed, view the log in **/var/log/Bootstrap** on the cluster VM. ● The scale-in operation cannot be rolled back. Therefore, the **Action upon Failure** can only be set to **Continue** after scale-in.

NOTE

The automation script is triggered only during auto scaling. It is not triggered when the cluster node is manually scaled out or in.

3.8 Managing Data Connections

3.8.1 Configuring Data Connections

MRS data connections are used to manage external source connections used by components in a cluster. For example, if Hive metadata uses an external relational database, a data connection can be used to associate the external relational database with the Hive component.

- **Local**: Metadata is stored in the local GaussDB of a cluster. When the cluster is deleted, the metadata is also deleted. To retain the metadata, manually back up the metadata in the database in advance.
- **Data Connection**: Metadata is stored in the associated PostgreSQL or MySQL database of the RDS service in the same VPC and subnet as the current cluster. When the cluster is terminated, the metadata is not deleted. Multiple MRS clusters can share the metadata.

NOTE

When Hive metadata is switched between different clusters, MRS synchronizes only the permissions in the metadata database of the Hive component. The permission model on MRS is maintained on MRS Manager. Therefore, when Hive metadata is switched between clusters, the permissions of users or user groups cannot be automatically synchronized to MRS Manager of another cluster.

Performing Operations Before Data Connection

Step 1 Log in to the RDS console.

Step 2 Click the **Instance Management** tab and click the name of the RDS DB instance used by the MRS data connection.

Step 3 Click **Log In** in the upper right corner to log in to the instance as user **root**.

Step 4 On the home page of the instance, click **Create Database** to create a database.

Step 5 On the top of the page, choose **Account Management > User Management**.

NOTE

If the selected data connection is **RDS MySQL database**, ensure that the database user is user **root**. If the user is not **root**, perform **Step 5** to **Step 7**.

Step 6 Click **Create User** to create a non-root user.

Step 7 On the top of the page, choose **SQL Operations > SQL Query**, switch to the target database by database name, and run the following SQL statements to grant permissions to the database user. In the following statements, *${db_name}* and *${db_user}* indicate the name of the database to be connected to MRS and the name of the new user, respectively.

```
grant SELECT, INSERT on mysql.* to '${db_user}'@'%' with grant option;
grant all privileges on ${db_name}.* to '${db_user}'@'%' with grant option;
grant reload on *.* to '${db_user}'@'%' with grant option;
flush privileges;
```

Step 8 Create a data connection by referring to **Creating a Data Connection**.

----End

Creating a Data Connection

Step 1 Log in to the MRS management console, and choose **Data Connections** in the left navigation pane.

Step 2 Click **Create Data Connection**.

Step 3 Set parameters according to **Table 3-13**.

Table 3-13 Data connection parameters

Parameter	Description
Type	Type of an external source connection. ● RDS for MySQL database. Clusters of that supports Hive or Ranger can connect to this type of database.
Name	Name of a data connection.

Parameter	Description
RDS Instance	RDS database instance. This instance must be created in RDS before being referenced here, and the database must have been created. For details, see **Performing Operations Before Data Connection**. Click **View RDS Instance** to view the created instances. **NOTE** ● To ensure network communications between the cluster and the PostgreSQL database, you are advised to create the instance in the same VPC and subnet as the cluster. ● The inbound rule of the security group of the RDS instance must allow access of the instance to port 3306. To configure that, click the instance name on the RDS console to go to the instance management page. In **Connection Information** area, click the name of **Security Group**. On the page that is displayed, click the **Inbound Rules** tab, and click **Add Rule**. On the displayed dialog box, in **Protocol & Port** area, select **TCP** and enter port number **3306**. In **Source** area, enter the IP address of all nodes where the MetaStore instance of Hive resides. ● Currently, MRS supports **PostgreSQL9.5/PostgreSQL9.6** on RDS. ● Currently, MRS supports only **MySQL 5.7.***x* on RDS.
Database	Name of the database to be connected to.
Username	Username for logging in to the database to be connected.
Password	Password for logging in to the database to be connected.

NOTE

If the selected data connection is an **RDS MySQL** database, ensure that the database user is a **root** user. If the user is not **root**, perform operations by referring to **Performing Operations Before Data Connection**.

Step 4 Click **OK**.

----End

Editing a Data Connection

Step 1 Log in to the MRS management console, and choose **Data Connections** in the left navigation pane.

Step 2 In the **Operation** column of the data connection list, click **Edit** in the row where the data connection to be edited is located.

Step 3 Modify parameters according to **Table 3-13**.

If the selected data connection has been associated with a cluster, the configuration changes will be synchronized to the cluster.

----End

Deleting a Data Connection

Step 1 Log in to the MRS management console, and choose **Data Connections** in the left navigation pane.

Step 2 In the **Operation** column of the data connection list, click **Delete** in the row where the data connection to be deleted is located.

If the selected data connection has been associated with a cluster, the deletion does not affect the cluster.

----End

Configuring a data connection during cluster creation

Step 1 Log in to the MRS console.

Step 2 Click **Create Cluster**. The **Create Cluster** page is displayed.

Step 3 Click the **Custom Config** tab.

Step 4 In the software configuration area, set **Metadata** by referring to **Table 3-14**. For other parameters, see **Creating a Custom Cluster** for configuration and cluster creation.

Table 3-14 Data connection parameters

Parameter	Description
Metadata	Whether to use external data sources to store metadata. ● **Local**: Metadata is stored in the local cluster. ● **Data connections**: Metadata of external data sources is used. If the cluster is abnormal or deleted, metadata is not affected. This mode applies to scenarios where storage and compute are decoupled. Clusters that support the Hive or Ranger component support this function.

Parameter	Description
Name	This parameter is available only when **Data connections** is selected for **Metadata**. It indicates the name of the component for which an external data source can be configured. ● Hive ● Ranger
Data Connection Type	This parameter is available only when **Data connections** is selected for **Metadata**. It indicates the type of an external data source. ● Hive supports the following data connection types: – RDS MySQL database – Local database ● Ranger supports the following data connection types: – RDS MySQL database – Local database
Data Connection Instance	This parameter is valid only when **Data Connection Type** is set to **RDS PostgreSQL database** or **RDS MySQL database**. This parameter indicates the name of the connection between the MRS cluster and the RDS database. This instance must be created before being referenced here. You can click **Create Data Connection** to create a data connection. For details, see **Performing Operations Before Data Connection** and **Creating a Data Connection**.

----End

3.8.2 Configuring Ranger Data Connections

Switch the Ranger metadata of the existing cluster to the metadata stored in the RDS database. This operation enables multiple MRS clusters to share the same metadata, and the metadata will not be deleted when the clusters are deleted. In this way, Ranger metadata migration is not required during cluster migration.

Prerequisites

You have created an RDS MySQL database instance. For details, see **Creating a Data Connection**.

NOTE

- For versions earlier than MRS 3.x, if the selected data connection is an **RDS MySQL database**, ensure that the database user is a **root** user. If the user is not **root**, create a user and grant permissions to the user by referring to **Performing Operations Before Data Connection**.
- In MRS 3.x or later, if the selected data connection is **RDS MySQL database**, the database user cannot be user **root**. In this case, create a user and grant permissions to the user by following the instructions provided in **Performing Operations Before Data Connection**.

Preparing for MySQL Database Ranger Metadata Configuration

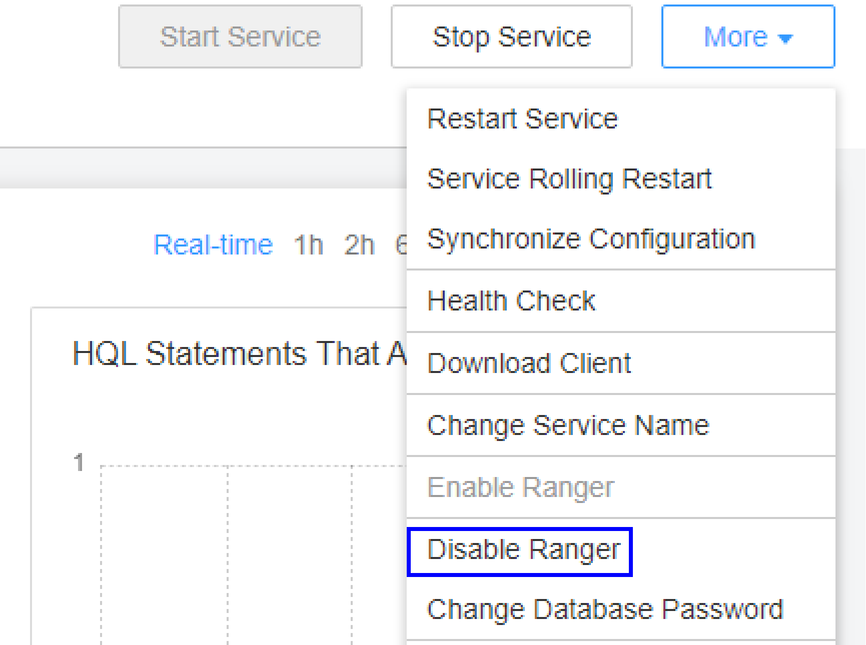
This operation is required only for **MRS 3.1.0 or later**.

Step 1 Log in to FusionInsight Manager. For details, see **Accessing FusionInsight Manager (MRS 3.x or Later)**. Choose **Clusters** > **Services** > *Service name*.

Currently, the following components in an MRS 3.1.*x* cluster support Ranger authentication: HDFS, HBase, Hive, Spark, Impala, Storm, and Kafka.

Step 2 In the upper right corner of the **Dashboard** page, click **More** and select **Disable Ranger**. If **Disable Ranger** is dimmed, Ranger authentication is disabled, as shown in **Figure 3-4**.

Figure 3-4 Disabling Ranger authentication



Step 3 (Optional) To use an existing authentication policy, perform this step to export the authentication policy on the Ranger web page. After the Ranger metadata is switched, you can import the existing authentication policy again. The following uses Hive as an example. After the export, a policy file in JSON format is generated in a local directory.

1. Log in to FusionInsight Manager.
2. Choose **Cluster** > **Services** > **Ranger** to go to the Ranger service overview page.

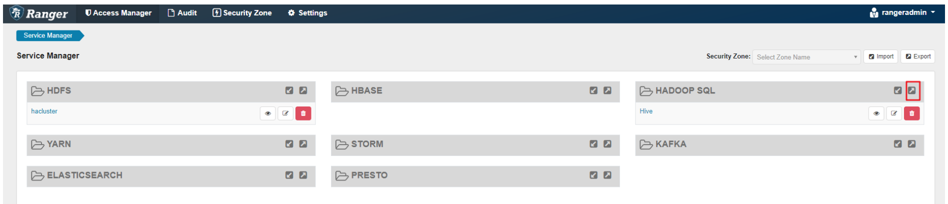
3. Click **RangerAdmin** in the **Basic Information** area to go to the Ranger web UI.

   The **admin** user in Ranger belongs to the **User** type. To view all management pages, click the username in the upper right corner and select **Log Out** to log out of the system.

4. Log in to the system as user **rangeradmin** (default password: **Rangeradmin@123**) or another user who has the Ranger administrator permissions.

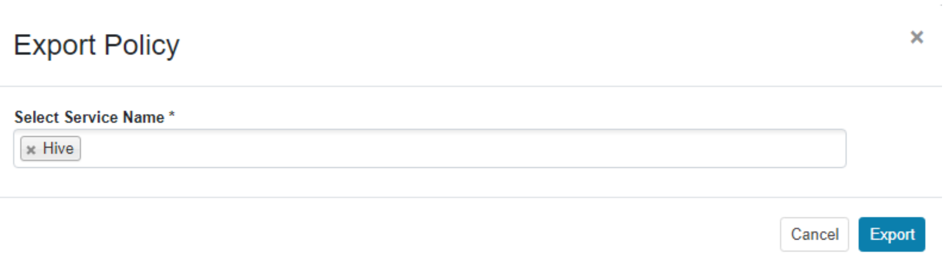
5. Click the export button in the row where the Hive component is located to export the authentication policy.

   Figure 3-5 Exporting authentication policies



6. Click **Export**. After the export is complete, a policy file in JSON format is generated in a local directory.

   Figure 3-6 Exporting Hive authentication policies



----End

Configuring a Data Connection for an MRS Cluster

Step 1 Log in to the MRS console.

Step 2 Click the name of the cluster to view its details.

Step 3 Click **Manage** on the right of **Data Connection** to go to the data connection configuration page.

Step 4 Click **Configure Data Connection** and set related parameters.

- **Component Name**: Ranger
- **Module Type**: Ranger metadata
- **Connection Type**: RDS MySQL database
- **Connection Instance**: Select a created RDS MySQL DB instance. To create a new data connection, see **Creating a Data Connection**.

Step 5 Select **I understand the consequences of performing the scale-in operation** and click **Test**.

Step 6 After the test is successful, click **OK** to complete the data connection configuration.

Step 7 Log in to FusionInsight Manager.

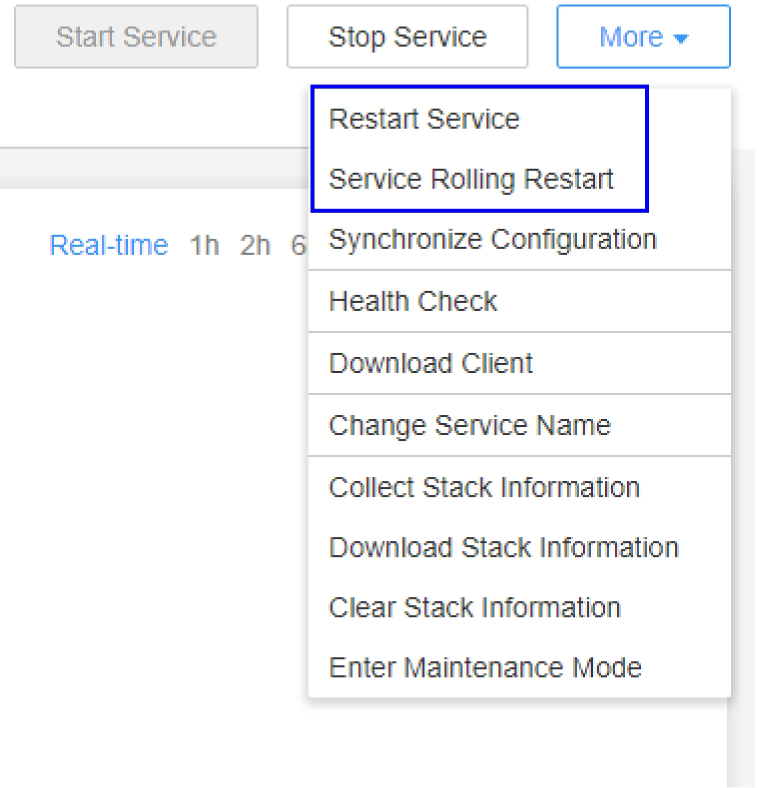
Step 8 Choose **Cluster** > **Services** > **Ranger** to go to the Ranger service overview page.

Step 9 Choose **More** > **Restart Service** or **More** > **Service Rolling Restart**.

If you choose **Restart Service**, services will be interrupted during the restart. If you select **Service Rolling Restart**, rolling restart can minimize the impact or do not affect service running.

Restarting Ranger will affect the permissions of all components controlled by Ranger and may affect the normal running of services. Therefore, restart Ranger when the cluster is idle or during off-peak hours. Before the Ranger component is restarted, the policies in the Ranger component still take effect.

Figure 3-7 Restarting a service

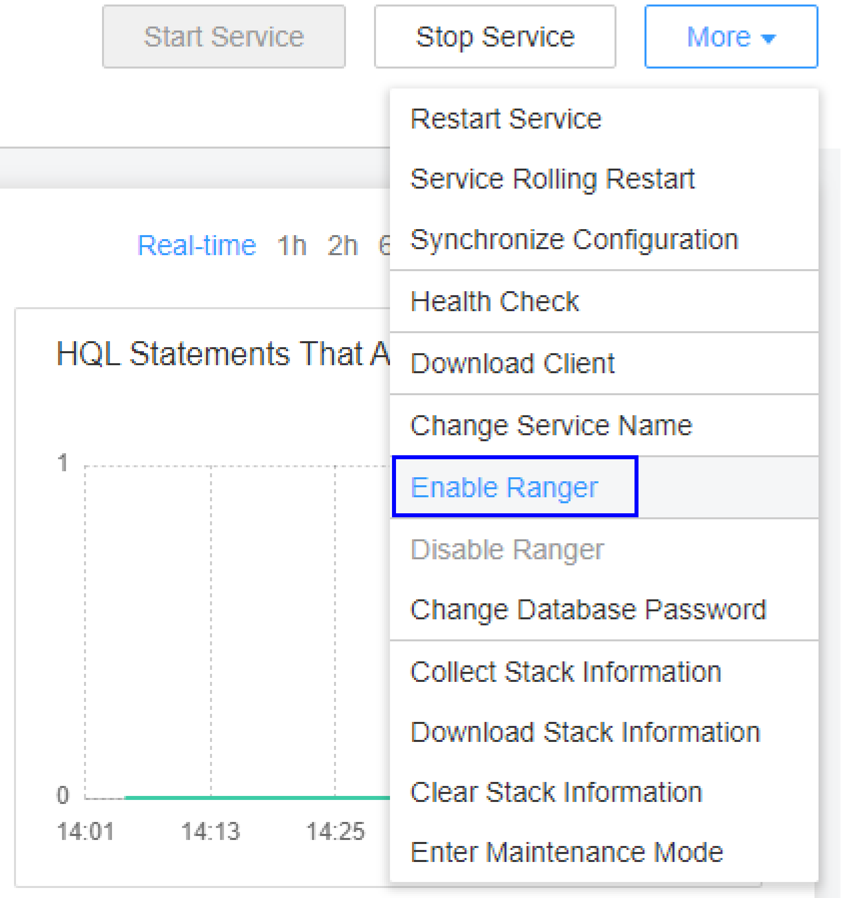


Step 10 Enable Ranger authentication for the component to be authenticated. The Hive component is used as an example.

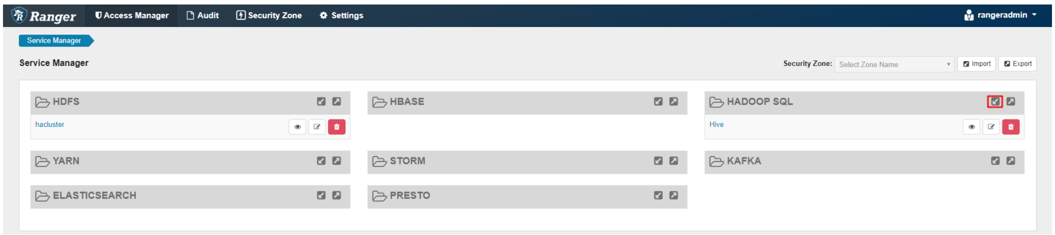
Currently, the following components in an MRS 3.1.*x* cluster support Ranger authentication: HDFS, HBase, Hive, Spark, Impala, Storm, and Kafka.

1. Log in to FusionInsight Manager and choose **Cluster** > **Services** > *Service Name*.
2. In the upper right corner of the **Dashboard** page, click **More** and select **Enable Ranger**.

Figure 3-8 Enabling Ranger authentication



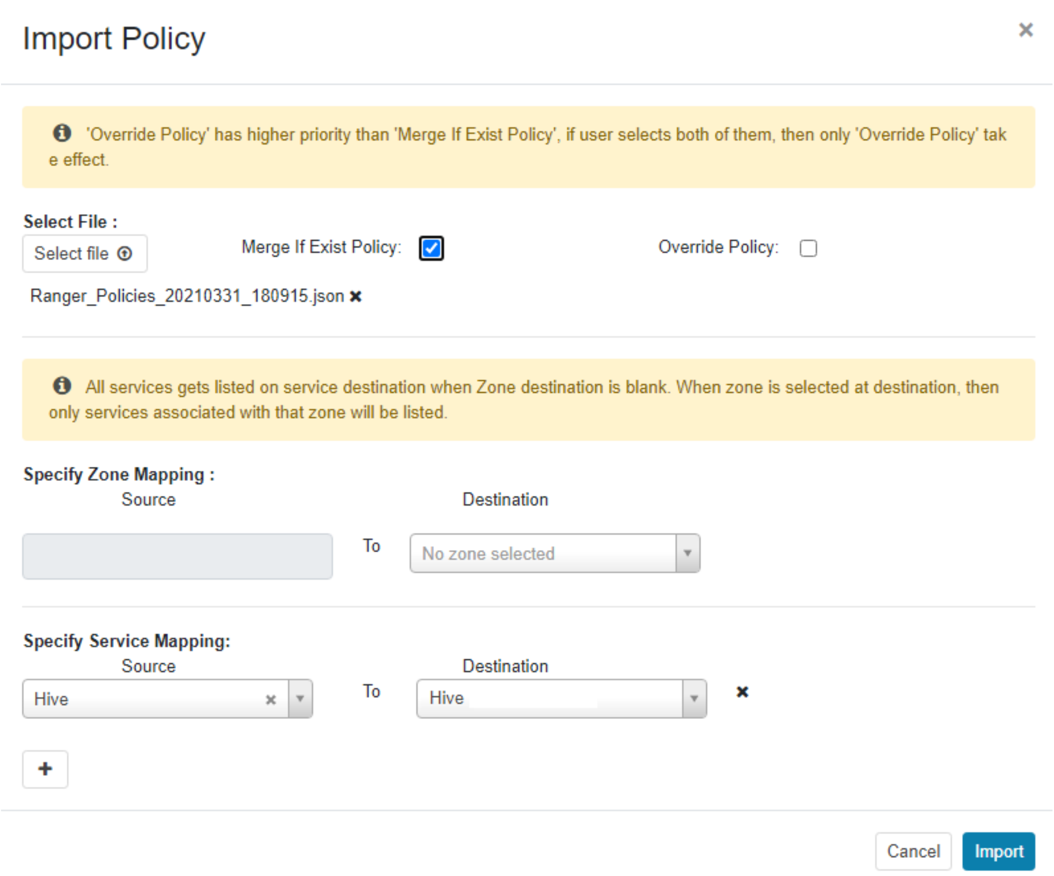
Step 11 Log in to the Ranger web UI and click the import button ☑ in the row of the Hive component.



Step 12 Import parameters.

- Click **Select file** and select the authentication policy file downloaded in **Step 3.6**.
- Select **Merge If Exist Policy**.

Figure 3-9 Importing authentication policies



Import Policy

'Override Policy' has higher priority than 'Merge If Exist Policy', if user selects both of them, then only 'Override Policy' tak e effect.

Select File : Select file

Merge If Exist Policy: ☑

Override Policy: □

Ranger_Policies_20210331_180915.json

All services gets listed on service destination when Zone destination is blank. When zone is selected at destination, then only services associated with that zone will be listed.

Specify Zone Mapping :

Source		Destination
	To	No zone selected

Specify Service Mapping:

Source		Destination
Hive	To	Hive

Step 13 Restart the component for which Ranger authentication is enabled.

1. Log in to FusionInsight Manager.
2. Choose **Cluster** > **Services** > **Hive** to go to the Hive service overview page.
3. Choose **More** > **Restart Service** or **More** > **Service Rolling Restart**.

   If you choose **Restart Service**, services will be interrupted during the restart. If you select **Service Rolling Restart**, rolling restart can minimize the impact or do not affect service running.

----End

3.8.3 Configuring a Hive Data Connection

This section describes how to switch the Hive metadata of an active cluster to the metadata stored in a local database or RDS database after you create a cluster. This operation enables multiple MRS clusters to share the same metadata, and the metadata will not be deleted when the clusters are deleted. In this way, Hive metadata migration is not required during cluster migration.

NOTE

- When Hive metadata is switched between different clusters, MRS synchronizes only the permissions in the metadata database of the Hive component. The permission model on MRS is maintained on MRS Manager. Therefore, when Hive metadata is switched between clusters, the permissions of users or user groups cannot be automatically synchronized to MRS Manager of another cluster.
- For clusters whose version is earlier than MRS 3.*x*, if the selected data connection is **RDS MySQL database**, ensure that the database user is **root**. If the user is not **root**, create a user and grant permissions to the user by referring to **Performing Operations Before Data Connection**.
- For clusters whose version is MRS 3.*x* or later, if the selected data connection is **RDS MySQL database**, the database user cannot be user **root**. In this case, create a user and grant permissions to the user by following the instructions provided in **Performing Operations Before Data Connection**.

Configuring a Hive Data Connection

This function is not supported in MRS 3.0.5.

Step 1 Log in to the MRS console. In the navigation pane on the left, choose **Clusters** > **Active Clusters**.

Step 2 Click the name of a cluster to go to the cluster details page.

Step 3 On the **Dashboard** tab page, click **Manage** next to **Data Connection**.

Step 4 On the **Data Connection** dialog box, the data connections associated with the cluster are displayed. You can click **Edit** or **Delete** to edit or delete the data connections.

Step 5 If there is no associated data connection on the **Data Connection** dialog box, click **Configure Data Connection** to add a connection.

NOTE

Only one data connection can be configured for a module type. For example, after a data connection is configured for Hive metadata, no other data connection can be configured for it. If no module type is available, the **Configure Data Connection** button is unavailable.

Table 3-15 Configuring a Hive data connection

Parameter	Description
Component	Hive
Module Type	Hive metadata
Data Connection Type	● RDS MySQL database ● Local database

Parameter	Description
Instance	This parameter is valid only when **Data Connection Type** is set to **RDS PostgreSQL database** or **RDS MySQL database**. Select the name of the connection between the MRS cluster and the RDS database. This instance must be created before being referenced here. You can click **Create Data Connection** to create a data connection. For details, see **Creating a Data Connection**.

Step 6 Click **Test** to test connectivity of the data connection.

Step 7 After the data connection is successful, click **OK**.

NOTE

- After Hive metadata is configured, restart Hive. Hive will create necessary database tables in the specified database. (If tables already exist, they will not be created.)
- Before restarting the Hive service, ensure that the driver package has been installed on all nodes where Metastore instances are located.
  - Postgres: Use the open source Postgres driver package to replace the existing one of the cluster. Upload the Postgres driver package **postgresql-42.2.5.jar** to the *${BIGDATA_HOME}***/third_lib/Hive** directory on all MetaStore instance nodes. To download the open-source driver package, visit **https://repo1.maven.org/maven2/org/postgresql/postgresql/42.2.5/**.
  - MySQL: Go to the MySQL official website (https://www.mysql.com/). Choose **DOWNLOADS** and click **MySQL Community (GPL) Downloads**. On the displayed page, click **Connector/J** to download the driver package of the corresponding version and upload the driver package to the **/opt/Bigdata/FusionInsight_HD_*/install/FusionInsight-Hive-*/hive-*/lib/** directory on all RDSMetastore nodes.

----End

3.9 Installing Third-Party Software Using Bootstrap Actions

This operation applies to MRS 3.*x* or earlier clusters.

In MRS 3.*x*, bootstrap actions cannot be added during cluster creation.

Prerequisites

The bootstrap action script has been prepared by referring to **Preparing the Bootstrap Action Script**.

Adding a Bootstrap Action When Creating a Cluster

Step 1 Log in to the MRS management console.

Step 2 Click **Create Cluster**. The page for creating a cluster is displayed.

Step 3 Click the **Custom Config** tab.

Step 4 Configure the cluster software and hardware by referring to **Creating a Custom Cluster**.

Step 5 On the **Set Advanced Options** tab page, click **Add** in the **Bootstrap Action** area.

Table 3-16 Parameters

Parameter	Description
Name	Name of a bootstrap action script The value can contain only digits, letters, spaces, hyphens (-), and underscores (_) and must not start with a space. The value can contain 1 to 64 characters. **NOTE** A name must be unique in the same cluster. You can set the same name for different clusters.
Script Path	Script path. The value can be an OBS file system path or a local VM path. ● An OBS file system path must start with **s3a://** and end with **.sh**, for example, **s3a://mrs-samples/*xxx*.sh**. ● A local VM path must start with a slash (/) and end with **.sh**. **NOTE** A path must be unique in the same cluster, but can be the same for different clusters.
Parameter	Bootstrap action script parameters
Execution Node	Select a type of the node where the bootstrap action script is executed.
Executed	Select the time when the bootstrap action script is executed. ● Before initial component start ● After initial component start
Action upon Failure	Whether to continue to execute subsequent scripts and create a cluster after the script fails to be executed. **NOTE** You are advised to set this parameter to **Continue** in the debugging phase so that the cluster can continue to be installed and started no matter whether the bootstrap action is successful.

Step 6 Click **OK**.

After the bootstrap action is added, you can edit, clone, or delete it in the **Operation** column.

----End

Adding an Automation Script on the Auto Scaling Page

Step 1 Log in to the MRS management console.

Step 2 Choose **Clusters** > **Active Clusters**, select a running cluster, and click its name to go to its details page.

Step 3 Click the **Nodes** tab. On this tab page, click **Auto Scaling** in the **Operation** column of the task node group. The **Auto Scaling** page is displayed.

If no task nodes are available, click **Configure Task Node** to add a task node and then perform this step.

NOTE

Configure Task Node is available only for MRS 3.*x* or later analysis, streaming, and hybrid clusters.

Step 4 Configure a resource plan.

Configuration procedure:

1. On the **Auto Scaling** page, enable **Auto Scaling**.
2. For example, the **Default Range** of node quantity is set to **2-2**, indicating that the number of task nodes is fixed to 2 except the time range specified in the resource plan.
3. Click **Configure Node Range for Specific Time Range** under **Default Range**.
4. Configure **Time Range** and **Node Range**. For example, set **Time Range** to **07:00-13:00**, and **Node Range** to **5-5**. This indicates that the number of task nodes is fixed to 5 in the time range specified in the resource plan. For details about the parameters, see **Table 3-11**.

   You can click **Configure Node Range for Specific Time Range** to configure multiple resource plans.

Step 5 (Optional) Configure automation scripts.

1. Set **Advanced Settings** to **Configure**.
2. Click **Create**. The **Automation Script** page is displayed.
3. Set **Name**, **Script Path**, **Execution Node**, **Parameter**, **Executed**, and **Action upon Failure**. For details about the parameters, see **Table 3-12**.
4. Click **OK** to save the automation script configurations.

Step 6 Select **I agree to authorize MRS to scale out or scale in nodes based on the above rule**.

Step 7 Click **OK**.

----End

3.10 Viewing Failed MRS Tasks

This section describes how to view and delete a failed MRS task.

Background

If a cluster fails to be created, terminated, scaled out, or scaled in, the **Manage Failed Tasks** page is displayed. Only the tasks that fail to be deleted are displayed on the **Cluster History** page. You can delete a failed task that is not required.

Procedure

Step 1 Log in to the MRS console.

Step 2 In the left navigation pane, choose **Clusters** > **Active Clusters**.

Step 3 Click ● or the number on the right of **Failed Tasks**. The **Manage Failed Tasks** page is displayed.

Step 4 In the **Operation** column of the cluster that you want to start, click **Delete**.

In this step, only one job can be deleted.

Step 5 You can click **Delete All** in the upper left corner of the task list to delete all failed tasks.

----End

3.11 Viewing Information of a Historical Cluster

Choose **Clusters > Cluster History** and click the name of a target cluster. You can view the cluster configuration and deployed node information.

The following table describes the parameters for the historical cluster information.

Table 3-17 Basic cluster information

Parameter	Description
Cluster Name	Name of a cluster. The cluster name is set when the cluster is created.
Cluster Status	Status of a cluster.
Cluster Version	Cluster version
Cluster Type	Type of the cluster to be created.
Obtaining a cluster ID	Unique identifier of a cluster, which is automatically assigned when a cluster is created
Created	Time when a cluster is created.
AZ	Availability zone (AZ) in the region of a cluster, which is set when a cluster is created.
Default Subnet	Subnet selected during cluster creation. A subnet provides dedicated network resources that are isolated from other networks, improving network security.

Parameter	Description
VPC	VPC selected during cluster creation. A VPC is a secure, isolated, and logical network environment.
OBS Permission Control	Click **Manage** and modify the mapping between MRS users and OBS permissions. For details, see **Configuring Fine-Grained Permissions for MRS Multi-User Access to OBS**.
Creating a data connection	Click **Manage** to view the data connection type associated with the cluster. For details, see **Configuring Data Connections**.
Agency	Click **Manage Agency** to bind or modify an agency for the cluster. An agency allows ECS or BMS to manage MRS resources. You can configure an agency of the ECS type to automatically obtain the AK/SK to access OBS. For details, see **Configuring a Storage-Compute Decoupled Cluster (Agency)**. The **MRS_ECS_DEFAULT_AGENCY** agency has the OBSOperateAccess permission of OBS and the CESFullAccess (for users who have enabled fine-grained policies), CES Administrator, and KMS Administrator permissions in the region where the cluster is located.
Key Pair	Name of a key pair. Set this parameter when creating a cluster. If the login mode is set to password during cluster creation, this parameter is not displayed.
Kerberos Authentication	Whether to enable Kerberos authentication when logging in to Manager. **NOTE** Kerberos authentication cannot be manually enabled or disabled after the cluster is created. Set this parameter with caution when creating a cluster. If you need to change the authentication status, you are advised to create a new cluster.
Security Group	Security group name of the cluster.
Data Disk Key Name	Name of the key used to encrypt data disks. To manage the used keys, log in to the key management console.
Data Disk Key ID	ID of the key used to encrypt data disks.
Component Version	Version of each component installed in the cluster.
License Version	License version of the cluster.
Agency	Delegates ECSs or BMSs to manage some of your resources.

Go back to the historical clusters page. You can use the following buttons to perform operations. For details about the buttons, see the following table.

Table 3-18 Icon description

Icon	Description
↻	Click ↻ to manually refresh the node information.
🔍	Enter a cluster name in the search bar and click 🔍 to search for a cluster.

4 Managing Clusters

4.1 Logging In to a Cluster

4.1.1 MRS Cluster Node Overview

This section describes remote login, MRS cluster node types, and node functions.

MRS cluster nodes support remote login. The following remote login methods are available:

- GUI login: Use the remote login function provided by the ECS management console to log in to the Linux interface of the Master node in the cluster.
- SSH login: Applies to Linux ECSs only. You can use a remote login tool (such as PuTTY) to log in to an ECS. The ECS must have a bound EIP.

  For details about how to apply for and bind EIP for the Master node, see **Virtual Private Cloud** > **User Guide** > **Elastic IP** > **Assigning an EIP and Binding It to an ECS**.

  You can log in to a Linux ECS using either a key pair or password.

NOTICE

If you need to use a key pair to access a cluster node, you need to log in to the node as user **root**. For details, see **Logging In to an ECS Using a Key Pair (SSH)**.

In an MRS cluster, a node is an ECS. **Table 4-1** describes the node types and node functions.

Table 4-1 Cluster node types

Node Type	Functions
Master node	Management node of an MRS cluster. It manages and monitors the cluster. In the navigation tree of the MRS management console, choose **Clusters** > **Active Clusters**, select a running cluster, and click its name to switch to the cluster details page. On the **Nodes** tab page, view the **Name**. The node that contains **master1** in its name is the Master1 node. The node that contains **master2** in its name is the Master2 node. You can log in to a Master node either using VNC on the ECS management console or using SSH. After logging in to the Master node, you can access Core nodes without entering passwords. The system automatically deploys the Master nodes in active/standby mode and supports the high availability (HA) feature for MRS cluster management. If the active management node fails, the standby management node switches to the active state and takes over services. To determine whether the Master1 node is the active management node, see **Determining Active and Standby Management Nodes of Manager**.
Core node	Work node of an MRS cluster. It processes and analyzes data and stores process data.
Task node	Compute node. It is used for auto scaling when the computing resources in a cluster are insufficient.

4.1.2 Logging In to an ECS

This section describes how to remotely log in to an ECS in an MRS cluster using the remote login (VNC mode) function provided on the ECS management console or a key or password (SSH mode). Remote login (VNC mode) is mainly used for emergency O&M. In other scenarios, it is recommended that you log in to the ECS using SSH.

NOTE

To log in to a cluster node using SSH, you need to manually add an inbound rule in the security group of the cluster. The source address is **Client IPv4 address/32** (or **Client IPv6 address/128**) and the port number is **22**. For details, see **Virtual Private Cloud** > **User Guide** > **Security > Security GroupAdding a Security Group Rule**.

Logging In to an ECS Using VNC

Step 1 Log in to the MRS management console.

Step 2 Choose **Clusters > Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.

Step 3 On the **Nodes** tab page, click the name of a Master node in the Master node group to log in to the ECS management console.

Step 4 In the upper right corner, click **Remote Login**.

Step 5 Perform subsequent operations by referring to **Login Using VNC**.

----End

Logging In to an ECS Using a Key Pair (SSH)

Logging In to the ECS from Local Windows

To log in to the Linux ECS from local Windows, perform the operations described in this section. The following procedure uses PuTTY as an example to log in to the ECS.

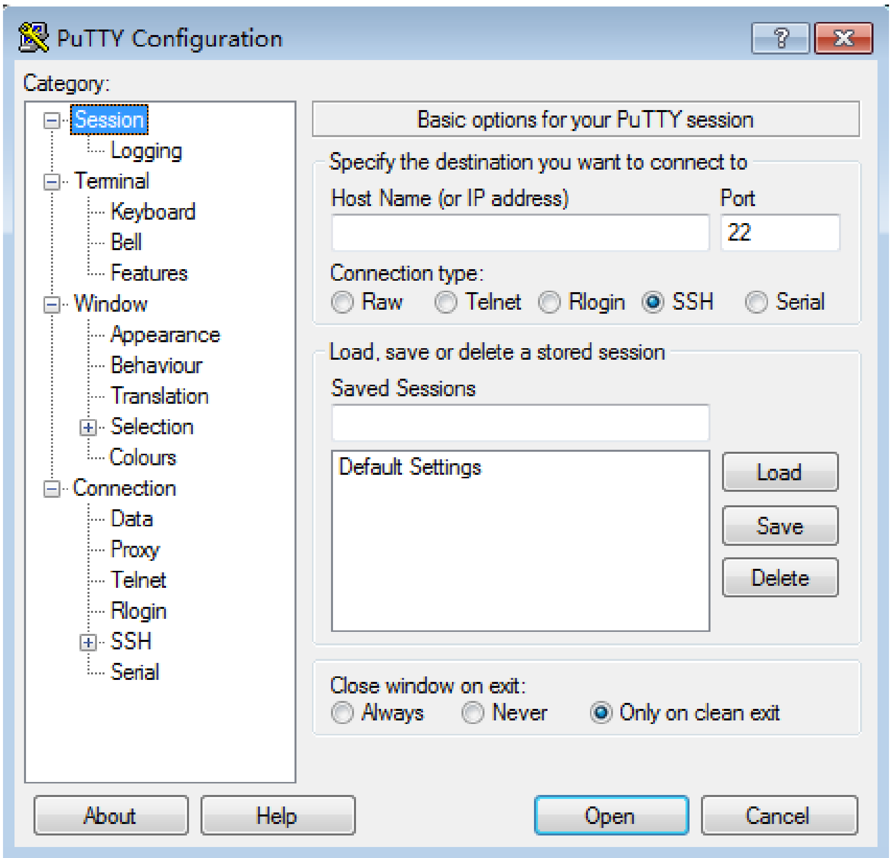
1. Log in to the MRS management console.
2. Choose **Clusters** > **Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.
3. On the **Nodes** tab page, click the name of a Master node in the Master node group to log in to the ECS management console.
4. Click the **EIPs** tab, click **Bind EIP** to bind an EIP to the ECS, and record the EIP. If an EIP has been bound to the ECS, skip this step.
5. Check whether the private key file has been converted to **.ppk** format.
   – If yes, go to **10**.
   – If no, go to **6**.
6. Run PuTTY.
7. In the **Actions** area, click **Load** and import the private key file you used during ECS creation.

   Ensure that the private key file is in the format of **All files (*.*)**.
8. Click **Save private key**.
9. Save the converted private key, for example, **kp-123.ppk**, to a local directory.
10. Run PuTTY.
11. Choose **Connection** > **Data**. Enter the image username in **Auto-login username**.

    NOTE

    The image username for cluster nodes is **root**.
12. Choose **Connection** > **SSH** > **Auth**. In the last configuration item **Private key file for authentication**, click **Browse** and select the private key converted in **9**.
13. Click **Session**.
    a. **Host Name (or IP address)**: Enter the EIP bound to the ECS.
    b. **Port**: Enter **22**.
    c. **Connection Type**: Select **SSH**.
    d. **Saved Sessions**: Task name, which can be clicked for remote connection when you use PuTTY next time

Figure 4-1 Clicking **Session**



Host Name (or IP address):
Port: 22

Connection type: ○ Raw ○ Telnet ○ Rlogin ● SSH ○ Serial

Saved Sessions: Default Settings

Close window on exit: ○ Always ○ Never ● Only on clean exit

14. Click **Open** to log in to the ECS.

    If you log in to the ECS for the first time, PuTTY displays a security warning dialog box, asking you whether to accept the ECS security certificate. Click **Yes** to save the certificate to your local registry.

Logging In to the ECS from Local Linux

To log in to the Linux ECS from local Linux, perform the operations described in this section. The following procedure uses private key file **kp-123.pem** as an example to log in to the ECS. The name of your private key file may differ.

1. On the Linux CLI, run the following command to change operation permissions:

   chmod 400 /*path*/kp-123.pem

   NOTE

   In the preceding command, *path* refers to the path where the key file is saved.

2. Run the following command to log in to the ECS:

   ssh -i /*path*/kp-123.pem *Default username@EIP*

   For example, if the default username is **root** and the EIP is **123.123.123.123**, run the following command:

   ssh -i /*path*/kp-123.pem root@123.123.123.123

NOTE

- *path* indicates the path where the key file is saved.
- *EIP* indicates the EIP bound to the ECS.
- The image username is **root** for cluster nodes.

4.1.3 Determining Active and Standby Management Nodes of Manager

This section describes how to determine the active and standby management nodes of Manager on the Master1 node.

Background

You can log in to other nodes in the cluster from the Master node. After logging in to the Master node, you can determine the active and standby management nodes of Manager and run commands on corresponding management nodes.

In active/standby mode, a switchover can be implemented between Master1 and Master2. For this reason, Master1 may not be the active management node for Manager.

Procedure

Step 1 Confirm the Master nodes of an MRS cluster.

1. In the navigation tree of the MRS management console, choose **Clusters > Active Clusters**, select a running cluster, and click its name to switch to the cluster details page. View basic information of the specified cluster.
2. On the **Nodes** tab page, view the node name. The node that contains **master1** in its name is the Master1 node. The node that contains **master2** in its name is the Master2 node.

Step 2 Determine the active and standby Manager management nodes.

1. Remotely log in to the Master1 node. For details, see **Logging In to an ECS**.
2. Run the following commands to switch the user:

   sudo su - root

   su - omm
3. Run the following command to identify the active and standby management nodes:

   For versions earlier than MRS 3.*x*, run the **sh ${BIGDATA_HOME}/om-0.0.1/sbin/status-oms.sh** command.

   For MRS 3.*x* or later: Run the **sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh** command.

   In the command output, the node whose **HAActive** is **active** is the active management node (mgtomsdat-sh-3-01-1 in the following example), and the node whose **HAActive** is **standby** is the standby management node (mgtomsdat-sh-3-01-2 in the following example).

```
Ha mode
double
NodeName          HostName               HAVersion       StartTime            HAActive
```

```
HAAllResOK          HARunPhase
192-168-0-30         mgtomsdat-sh-3-01-1          V100R001C01        2014-11-18 23:43:02
active             normal              Actived
192-168-0-24         mgtomsdat-sh-3-01-2          V100R001C01        2014-11-21 07:14:02
standby             normal              Deactived
```

NOTE

If the Master1 node to which you have logged in is the standby management node and you need to log in to the active management node, run the following command:

ssh *IP address of Master2 node*

----End

4.2 Cluster Overview

4.2.1 Cluster List

You can quickly view the status of all clusters and jobs by viewing the dashboard information, and obtain relevant MRS documents from in the left navigation pane on the MRS console.

MRS is used to manage and analyze massive data. It is easy to use. You can create a cluster and add MapReduce, Spark, and Hive jobs to the cluster to analyze and process user data. After being processed, you can transmit the data in SSL encryption mode to OBS to ensure data integrity and confidentiality.

Cluster Status

Table 4-2 lists the statuses of all MRS clusters after you log in to the MRS management console.

Table 4-2 Cluster status

Status	Description
Starting	If a cluster is being created, the cluster is in the **Starting** state.
Running	If a cluster is created successfully and all components in the cluster are normal, the cluster is in the **Running** state.
Scaling out	If the Core or Task node in a cluster is being added, the cluster is in the **Scaling out** state. **NOTE** If the cluster scale-out fails, you can add node to the cluster again.
Scaling in	If you stop, delete, change or reinstall the OSs of cluster nodes, and modify the specifications of the cluster node, the cluster nodes are being terminated. Then, the cluster is in the **Scaling in** state.
Abnormal	If some components in a cluster are abnormal, the cluster is **Abnormal**.

Status	Description
Terminating	If a cluster node is being terminated, the cluster is in the **Terminating** state.
Terminated	The cluster has been terminated. This parameter is displayed only in **Cluster History**.
Scaling up Master node	If the specifications of a master node are being upgraded, the cluster status is **Scaling up Master node**.

Job Status

Table 4-3 describes the status of jobs that you execute after logging in to the MRS management console.

Table 4-3 Job status

Status	Description
Accepted	Initial status of a job after it is successfully submitted.
Running	A job is being executed.
Completed	A job has been executed and completed successfully.
Terminated	A job is stopped during execution.
Abnormal	An error occurs during job execution or job execution fails.

4.2.2 Checking the Cluster Status

The cluster list contains all clusters in MRS. You can view clusters in various states. If a large number of clusters are involved, navigate through multiple pages to view all of the clusters.

MRS, as a platform managing and analyzing massive data, provides a PB-level data processing capability. MRS allows you to create multiple clusters. The cluster quantity is subject to that of ECSs.

Clusters are listed in chronological order by default in the cluster list, with the most recent cluster displayed at the top. **Table 4-4** describes the cluster list parameters.

- **Active Clusters**: contain all clusters except the clusters in the **Failed** and **Terminated** states.
- **Cluster History**: contains the tasks in the **Terminated** states. Only clusters terminated within the last six months are displayed. If you want to view clusters terminated six months ago, contact technical support engineers.
- **Failed Tasks**: only contain the tasks in the **Failed** state. Task failures include:
  - Cluster creation failure

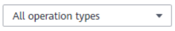
- Cluster termination failure
- Cluster scale-out failure
- Cluster scale-in failure
- Cluster patch installation failure (supported only by versions earlier than MRS 3.x)
- Cluster patch uninstallation failure (supported only by versions earlier than MRS 3.x)
- Cluster specifications upgrade failure

Table 4-4 Parameters in the active cluster list

Parameter	Description
Name/ID	Cluster name, which is set when a cluster is created. Unique identifier of a cluster, which is automatically assigned when a cluster is created. ● ✎: Change the cluster name. ● ⧉: Copy the cluster ID.
Cluster Version	Cluster version.
Nodes	Number of nodes that can be deployed in a cluster. This parameter is set when a cluster is created.

Parameter	Description
Status	Status and operation progress description of a cluster. The cluster creation progress includes: ● Verifying cluster parameters ● Applying for cluster resources ● Creating VMs ● Initializing VMs ● Installing MRS Manager ● Deploying the cluster ● Cluster installation failed The cluster scale-out progress includes: ● Preparing for scale-out ● Creating VMs ● Initializing VMs ● Adding nodes to the cluster ● Scale-out failed The cluster scale-in progress includes: ● Preparing for scale-in ● Decommissioning instance ● Deleting VMs ● Deleting nodes from the cluster ● Scale-in failed The system will display causes of cluster installation, scale-out, and scale-in failures. For details, see **Table 3-5**.
Terminated	Time when a cluster node stops and the cluster node begins to be terminated. This parameter is valid only for historical clusters displayed on the **Cluster History** page.
AZ	Availability zone (AZ) in the region of a cluster, which is set when a cluster is created.
Operation	**Terminate**: If you want to terminate a cluster after jobs are complete, click **Terminate**. The cluster status changes from **Running** to **Terminating**. After the cluster is terminated, the cluster status will change to **Terminated** and will be displayed in **Cluster History**. If the MRS cluster fails to be deployed, the cluster is automatically terminated. This parameter is displayed in **Active Clusters** only. **NOTE** Typically after data is analyzed and stored, or when the cluster encounters an exception and cannot work, you can terminate a cluster. If a cluster is terminated before data processing and analysis are completed, data loss may occur. Therefore, exercise caution when terminating a cluster.

Table 4-5 Button description

Button	Description
 All operation types	In the drop-down list, select a status to filter clusters: ● Active Clusters – All operation types: displays all existing clusters. – Starting: displays existing clusters in the **Starting** state. – Running: displays existing clusters in the **Running** state. – Scaling out: displays existing clusters in the **Scaling out** state. – Scaling in: displays existing clusters in the **Scaling in** state. – Abnormal: displays existing clusters in the **Abnormal** state. – Terminating: displays existing clusters in the **Terminating** state.
●	Choose **Clusters** > **Active Clusters** and click ● to go to the page for managing failed tasks. ● *Num*: displays the failed tasks in the **failed** state.
🔍	Enter a cluster name in the search bar and click 🔍 to search for a cluster.
Search by Tag	Click **Search by Tag**, enter the tag of the cluster to be queried, and click **Search** to search for the clusters. You can select a tag key or tag value from their drop-down lists. When the tag key or tag value is exactly matched, the system can automatically locate the target cluster. If you enter multiple tags, their intersections are used to search for the cluster.
↻	Click ↻ to manually refresh the cluster list.

4.2.3 Viewing Basic Cluster Information

You can monitor and manage the clusters you have created. Choose **Clusters** > **Active Clusters**. Select a cluster and click its name to go to the cluster details page. On the displayed page, view the basic configuration and node information of the cluster.

NOTE

On the MRS console, operations performed on an ECS cluster are basically the same as those performed on a BMS cluster. This document describes operations on an ECS cluster. If operations on the two clusters differ, the operations will be described separately.

On the cluster details page, click **Dashboard**. **Table 4-6** describes the parameters on the **Dashboard** tab page.

Table 4-6 Basic cluster information

Parameter	Description
Cluster Name	Name of a cluster. Set this parameter when creating a cluster. Click ✎ to change the cluster name. For versions earlier than MRS 3.x, only the cluster name displayed on the MRS management console is changed, while the cluster name on MRS Manager is not changed synchronously.
Cluster Status	Cluster status. For details, see **Table 4-2**.
MRS Manager	Portal for the Manager page. ● For MRS 3.*x* or later, see **Accessing FusionInsight Manager (MRS 3.x or Later)**. ● For versions earlier than MRS 3.x, you need to bind an EIP and add a security group rule as prompted before accessing the MRS Manager page. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.
Cluster Version	MRS version information.
Cluster Type	There are three types of clusters: ● **Analysis cluster**: is used for offline data analysis and provides Hadoop components. ● **Streaming cluster**: is used for streaming tasks and provides stream processing components. ● **Hybrid cluster**: is used for both offline data analysis and streaming processing and provides Hadoop components and streaming processing components. ● Custom: An MRS cluster with all custom components. MRS 3.*x* and later versions support this type.
Cluster ID	Unique identifier of a cluster, which is automatically assigned when a cluster is created.
Created	Time when a cluster is created.
AZ	Availability zone (AZ) in the region of a cluster, which is set when a cluster is created.

Parameter	Description
Default Subnet	Subnet selected during cluster creation. If the subnet IP addresses are insufficient, click **Change Subnet** to switch to another subnet in the same VPC of the current cluster to obtain more available subnet IP addresses. Changing a subnet does not affect the IP addresses and subnets of existing nodes. A subnet provides dedicated network resources that are isolated from other networks, improving network security.
VPC	VPC selected during cluster creation. A VPC is a secure, isolated, and logical network environment.
Elastic IP (EIP)	After binding an EIP to an MRS cluster, you can use the EIP to access the Manager web UI of the cluster.
OBS Permission Control	Click **Manage** and modify the mapping between MRS users and OBS permissions. For details, see **Configuring Fine-Grained Permissions for MRS Multi-User Access to OBS**.
Data Connection	Click **Manage** to view the data connection type associated with the cluster. For details, see **Configuring Data Connections**.
Agency	Click **Manage Agency** to bind or modify an agency for the cluster. An agency allows ECS or BMS to manage MRS resources. You can configure an agency of the ECS type to automatically obtain the AK/SK to access OBS. For details, see **Configuring a Storage-Compute Decoupled Cluster (Agency)**. The **MRS_ECS_DEFAULT_AGENCY** agency has the OBS OperateAccess permission of OBS and the CES FullAccess (for users who have enabled fine-grained policies), CES Administrator, and KMS Administrator permissions in the region where the cluster is located.
Key Pair	Name of a key pair. Set this parameter when creating a cluster. If the login mode is set to password during cluster creation, this parameter is not displayed.
Kerberos Authentication	Whether to enable Kerberos authentication when logging in to Manager.
Logging	Used to collect logs about cluster creation and scaling failures.
Security Group	Security group name of the cluster.
Data Disk Key Name	Name of the key used to encrypt data disks. To manage the used keys, log in to the key management console.
Data Disk Key ID	ID of the key used to encrypt data disks.

Parameter	Description
IAM User Synchronization	IAM user information can be synchronized to an MRS cluster for cluster management. For details, see **Synchronizing IAM Users to MRS**. **NOTE** The **Components**, **Tenants**, and **Backups & Restorations** tab pages on the cluster details page can be used only after users are synchronized. After clusters of MRS 3.x are synchronized, you can use the **Component Management** function.
Secure Communications	Used to display the security authorization status. You can click to enable or disable security authorization. Disabling security authorization brings high risks. Exercise caution when performing this operation. For details, see **Communication Security Authorization**.

Table 4-7 Component versions

Parameter	Description
Hadoop Version	Displays the Hadoop version information.
Spark Version	Version of the Spark component. Only clusters of versions earlier than MRS 3.x support this parameter.
HBase Version	Displays the HBase version information.
Hive Version	Displays the Hive version information.
Hue Version	Displays the Hue version information.
Loader Version	Displays the Loader version information.
Kafka Version	Displays the Kafka version information.
Storm Version	Displays the Storm version information.
Flume Version	Displays the Flume version information.
Tez Version	Displays the Tez version information.
Presto Version	Displays the Presto version information.
KafkaManager Version	Displays the KafkaManager version information.
Flink Version	Displays the Flink version information.
Alluxio Version	Displays the Alluxio version information.
Ranger Version	Displays the Ranger version information.

Parameter	Description
Impala Version	Displays the Impala version information.
Kudu Version	Displays the Kudu version information.
Spark2x Version	Displays the version information about the Spark2x component. Only clusters of MRS 3.x or later support this function.
Oozie Version	Displays the Oozie version information. Only clusters of MRS 3.x or later support this function.
ClickHouse Version	Displays ClickHouse version information. Only clusters of MRS 3.x or later support this function.

On the cluster details page, click **Nodes**. For details about the node parameters, see **Table 4-8**.

Table 4-8 Node information

Parameter	Description
Configure Task Node	Used to add a Task node. For details, see **Adding a Task Node**. For 3.x and later versions, this operation applies only to the analysis cluster, streaming cluster, and hybrid cluster.
Add Node Group	This parameter applies only to 3.x and later versions. It applies to customized clusters only and is used to add node groups. For details, see **Adding a Node Group**.
Node Group	Node group name.
Node Type	Node type: ● **Master**: A Master node in an MRS cluster manages the cluster, assigns MapReduce executable files to Core nodes, traces the execution status of each job, and monitors the DataNode running status. ● A task node group is a group of nodes where only data roles that do not store data are deployed. The roles include NodeManager, ThriftServer, Thrift1Server, RESTServer, Supervisor, LogViewer, HBaseIndexer, and TagSync. ● If other roles are deployed in the node group in addition to the preceding roles, the node group is the Core node group. On the **Nodes** tab page, click ⌄ next to a node group name to unfold the nodes contained in the node group. For details about the parameters, see **Managing Components and Monitoring Hosts**.
Node Count	Number of nodes in a node group.

Parameter	Description
Operation	● **Scale Out**: For details, see **Manually Scaling Out a Cluster**. ● **Scale In**: For details, see **Manually Scaling In a Cluster**. ● **Auto Scaling**: For details, see **Configuring an Auto Scaling Rule**. ● **View Roles**: You can view information about roles deployed on the node group. This function applies only to custom clusters of 3.x and later.

4.2.4 Viewing Cluster Patch Information

To view patch information about cluster components, you can download the required patch if the cluster component, such as Hadoop or Spark, is faulty. On the MRS console, choose **Clusters > Active Clusters**, select a cluster, and click the cluster name. On the cluster details page that is displayed, upgrade the component and rectify the fault.

NOTE

MRS 3.x does not have patch version information. Therefore, this section is not involved.

- Patch Name: name of the patch package
- Published: time when the patch package is released
- Status: patch status
- Patch Description: patch version description
- Operation: patch installation or uninstallation

4.2.5 Viewing and Customizing Cluster Monitoring Metrics

MRS cluster nodes are classified into management nodes, control nodes, and data nodes. The change trends of key host monitoring metrics on each type of node can be calculated and displayed as curve charts in reports based on the customized periods. If a host belongs to multiple node types, the metric statistics will be repeatedly collected.

This section provides overview of MRS clusters and describes how to view, customize, and export node monitoring metrics on MRS Manager.

NOTE

Cluster metrics are monitored periodically. The average historical monitoring interval is about 5 minutes.

Method 1 (applicable to clusters of versions earlier than MRS 3.x):

Step 1 Choose **Clusters** > **Active Clusters** and click a cluster name to go to the cluster details page.

Step 2 Click the **Dashboard** tab, you can view the cluster host health status statistics on the lower part of the displayed tab page.

Step 3 To view or export reports of other metrics, click **Access Manager** next to **MRS Manager** in the **Basic Information** area to access the Manager page. For details, see **Accessing Manager**.

Step 4 On the Manager page, view, customize, and export the node monitoring metric report. For details, see **Dashboard**.

----End

Method 2 (applicable to clusters of versions earlier than MRS 3.x)

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters > Active Clusters** and click a cluster name to go to the cluster details page.

Step 3 In the **Basic Information** area on the **Dashboard** tab page, click **Click to synchronize** on the right side of **IAM User Sync** to synchronize IAM users.

Step 4 After the synchronization is complete, you can view the cluster monitoring metric report on the right of the page.

Step 5 In time range area, specify a period to view monitoring data. The options are as follows:

- Last 1 hour
- Last 3 hours
- Last 12 hours
- Last 24 hours
- Recent 7 days
- Recent 30 days
- Customize: You can customize the period for viewing monitoring data.

Step 6 Customize a monitoring metric report.

1. Click **Customize** and select monitoring metrics to be displayed.

   MRS supports a maximum of 14 monitoring metrics, but at most 12 customized monitoring metrics can be displayed on the page.

   - Cluster Host Health Status
   - Cluster Network Read Speed Statistics
   - Host Network Read Speed Distribution
   - Host Network Write Speed Distribution
   - Cluster Disk Write Speed Statistics
   - Cluster Disk Usage Statistics
   - Cluster Disk Information
   - Host Disk Usage Statistics
   - Cluster Disk Read Speed Statistics
   - Cluster Memory Usage Statistics
   - Host Memory Usage Distribution
   - Cluster Network Write Speed Statistics
   - Host CPU Usage Distribution

  - Cluster CPU Usage Statistics

2. Click **OK** to save the selected monitoring metrics for display.

   NOTE

   Click **Clear** to cancel all the selected monitoring metrics in a batch.

Step 7 Export a monitoring report.

1. Select a period. The options are as follows:
   - Last 1 hour
   - Last 3 hours
   - Last 12 hours
   - Last 24 hours
   - Recent 7 days
   - Recent 30 days
   - Customize: You can customize the period for viewing monitoring data.
2. Click **Export**. MRS will generate a report about the selected monitoring metrics in a specified time of period. Save the report.

----End

Method 3: (applicable to MRS 3.x clusters)

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters > Active Clusters** and click a cluster name to go to the cluster details page.

Step 3 In the **Basic Information** area on the **Dashboard** tab page, click **Click to synchronize** on the right side of **IAM User Sync** to synchronize IAM users.

Step 4 After the synchronization is complete, you can view the cluster monitoring metric report on the right of the page.

Step 5 In time range area, specify a period to view monitoring data. The options are as follows:

- Last 1 hour
- Last 3 hours
- Last 12 hours
- Last 24 hours
- Recent 7 days
- Recent 30 days
- Customize: You can customize the period for viewing monitoring data.

Step 6 Customize a monitoring metric report.

1. Click **Customize** and select monitoring metrics to be displayed.

   At most 12 customized monitoring metrics can be displayed on the page.
2. Click **OK** to save the selected monitoring metrics for display.

NOTE

Click **Clear** to cancel all the selected monitoring metrics in a batch.

----End

4.2.6 Managing Components and Monitoring Hosts

You can manage the following status and metrics of all components (including role instances) and hosts on the MRS console:

- Status information: includes operation, health, configuration, and role instance status.
- Indicator information: includes key monitoring indicators for each component.
- Export monitoring metrics. (This function is not supported in MRS 3.x or later.)

NOTE

- For versions earlier than MRS 3.x, see **Managing Services and Monitoring Hosts**.
- For MRS 3.x or later, see **Procedure**.
- You can set the interval for automatically refreshing the page or click ⟳ to refresh the page immediately.
- Component management supports the following parameter values:
  - Refresh every 30 seconds
  - Refresh every 60 seconds
  - Stop refreshing

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Managing Components

NOTE

For details about how to perform operations on MRS Manager, see **Managing Service Monitoring**.

Step 1 On the MRS cluster details page, click **Components**.

On the **Components** tab page, **Service**, **Operating Status**, **Health Status**, **Configuration Status**, **Role**, and **Operation** are displayed in the component list.

- **Table 4-9** describes the service operating status.

Table 4-9 Service operating status

Status	Description
Started	The service is started.

Status	Description
Stopped	The service is stopped.
Failed to start	Failed to start the role instance.
Failed to stop	Failed to stop the service.
Unknown	Indicates initial service status after the background system restarts.

- **Table 4-10** describes the service health status.

Table 4-10 Service health status

Status	Description
Good	Indicates that all role instances in the service are running properly.
Faulty	Indicates that the running status of at least one role instance is **Faulty** or the status of the service on which the current service depends is abnormal.
Unknown	Indicates that all role instances in the service are in the **Unknown** state.
Restoring	Indicates that the background system is restarting the service.
Partially Healthy	Indicates that the status of the service on which the service depends is abnormal, and APIs related to the abnormal service cannot be invoked by external systems.

- **Table 4-11** describes the service health status.

Table 4-11 Service configuration status

Status	Description
Synchronized	The latest configuration takes effect.
Configuration expired	The latest configuration does not take effect after the parameter modification. Related services need to be restarted.
Configuration failed	The communication is incorrect or data cannot be read or written during the parameter configuration. Use **Synchronize Configuration** to rectify the fault.
Configuring	Parameters are being configured.
Unknown	Indicates that configuration status cannot be obtained.

By default, the **Service** column is sorted in ascending order. You can click the icon next to **Service**, **Operating Status**, **Health Status**, or **Configuration Status** to change the sorting mode.

Step 2 Click a specified service in the list to view its status and metric information.

Step 3 Customize and view monitoring graphs.

1. In the **Charts** area, click **Customize** to customize service monitoring metrics.
2. In **Period** area, select a time of period and click **View** to view the monitoring data within the time period.

----End

Managing Role Instances

NOTE

For versions earlier than MRS 3.x, see **Managing Role Instances**.

Step 1 On the MRS cluster details page, click **Components**. In the component list, click the specified service name.

Step 2 Click **Instances** to view the role status.

The role instance list contains the Role, Host Name, Management IP Address, Service IP Address, Rack, Running Status, and Configuration Status of each instance.

- **Table 4-12** shows the running status of a role instance.

Table 4-12 Role instance running status

Status	Description
Good	Indicates that the instance is running properly.
Bad	Indicates that the instance cannot run properly.
Decommissioned	Indicates that the instance is out of service.
Not started	Indicates that the instance is stopped.
Unknown	Indicates that the initial status of the instance cannot be detected.
Starting	Indicates that the instance is being started.
Stopping	Indicates that the instance is being stopped.
Restoring	Indicates that an exception may occur in the instance and the instance is being automatically rectified.
Decommissioning	Indicates that the instance is being decommissioned.
Recommissioning	Indicates that the instance is being recommissioned.

Status	Description
Failed to start	Indicates that the service fails to be started.
Failed to stop	Indicates that the service fails to be stopped.

- Table 4-13 shows the configuration status of a role instance.

Table 4-13 Role instance configuration status

Status	Description
Synchronized	The latest configuration takes effect.
Configuration expired	The latest configuration does not take effect after the parameter modification. Related services need to be restarted.
Configuration failed	The communication is incorrect or data cannot be read or written during the parameter configuration. Use **Synchronize Configuration** to rectify the fault.
Configuring	Parameters are being configured.
Unknown	Current configuration status cannot be obtained.

By default, the **Role** column is sorted in ascending order. You can click the sorting icon next to **Role**, **Host Name**, **OM IP Address**, **Business IP Address**, **Rack**, **Running Status**, or **Configuration Status** to change the sorting mode.

You can filter out all instances of the same role in the **Role** column.

You can set search criteria in the role search area by clicking **Advanced Search**, and click **Search** to view specified role information. You can click **Reset** to reset the search criteria. Fuzzy search is supported.

Step 3 Click the target role instance to view its status and metric information.

Step 4 Customize and view monitoring graphs.

1. In the **Charts** area, click **Customize** to customize service monitoring metrics.
2. In **Period** area, select a time of period and click **View** to view the monitoring data within the time period.

----End

Managing Hosts

NOTE

For versions earlier than MRS 3.x, see Managing Hosts.

Step 1 On the MRS cluster details page, click the **Nodes** tab and expand a node group to view the host status.

The host list contains the **Node Name**, **IP Address**, **Rack**, **Operating Status**, **Health Status**, **CPU Usage**, **Memory Usage**, **Disk Usage**, **Network Speed**, Specification Name, **Specifications** and **AZ**.

- **Table 4-14** shows the host operating status.

Table 4-14 Host operating status

Status	Description
Normal	The host and service roles on the host are running properly.
Isolated	The host is isolated, and the service roles on the host stop running.

- **Table 4-15** describes the host health status.

Table 4-15 Host health status

Status	Description
Good	The host can properly send heartbeats.
Bad	The host fails to send heartbeats due to timeout.
Unknown	The host initial status is unknown during the operation of adding or deleting a host.

The nodes are sorted in ascending order by default. You can click **Node Name**, **IP Address**, **Rack**, **Operating Status**, **Health Status**, **CPU Usage**, **Memory Usage**, **Disk Usage**, **Network Speed**, **Specification Name**, or **Specifications** to change the sorting mode.

Step 2 Click the target node in the list to view its status and metric information.

----End

4.3 Cluster O&M

4.3.1 Importing and Exporting Data

Through the **Files** tab page, you can create, delete, import, export, delete files in the analysis cluster. Currently, file creation is not supported. Streaming clusters do not support the file management function on the MRS GUI. In a cluster with Kerberos authentication enabled, to read or write the folders in the root directory, add a role that has the required permissions on the folders by referring to **Creating a Role**. Then, add the new role to the user group to which the user who submits the job belongs by referring to **Related Tasks**.

Background

Data sources processed by MRS are from OBS or HDFS. OBS is an object-based storage service that provides you with massive, secure, reliable, and cost-effective data storage capabilities. MRS can process data in OBS directly. You can view, manage, and use data by using the web page of the management control platform or OBS client. In addition, you can use REST APIs independently or integrate APIs to service applications to manage and access data.

Before creating jobs, upload the local data to OBS for MRS to compute and analyze. MRS allows exporting data from OBS to HDFS for computing and analyzing. After the data analysis and computing are completed, you can store the data in HDFS or export them to OBS. HDFS and OBS can also store the compressed data in the format of **bz2** or **gz**.

Importing Data

Currently, MRS can only import data from OBS to HDFS. The file upload rate decreases with the increase of the file size. This mode applies to scenarios where the data volume is small.

You can perform the following steps to import files and directories:

1. Log in to the MRS console.
2. Choose **Clusters > Active Clusters** and click the name of the cluster to be queried to enter the page displaying the cluster's information.
3. Click the **Files** tab, and go to the file management page.
4. Select **HDFS File List**.
5. Go to the data storage directory, for example, **bd_app1**.

   The **bd_app1** directory is only an example. You can use any directory on the page or create a new one.

   The requirements for creating a folder are as follows:

   - The folder name contains a maximum of 255 characters
   - The folder name cannot be empty.
   - The folder name cannot contain the following special characters: /:*?"<>|\;&,'`!{}[]$%+
   - The value cannot start or end with a period (.).
   - The spaces at the beginning and end are ignored.
6. Click **Import Data** and configure the HDFS and OBS paths correctly. When configuring the OBS or HDFS path, click **Browse**, select a file directory, and click **Yes**.
   - OBS path
     - The path must start with **obs://**.
     - Files or programs encrypted by KMS cannot be imported.
     - An empty folder cannot be imported.
     - The directory and file name can contain letters, digits, hyphens (-), and underscores (_), but cannot contain the following special characters: ;|&>,<'$*?\

        - The directory and file name cannot start or end with a space, but can contain spaces between them.
        - The OBS full path contains a maximum of characters.
    - HDFS path
        - The path starts with **/user** by default.
        - The directory and file name can contain letters, digits, hyphens (-), and underscores (_), but cannot contain the following special characters: ;|&>,<'$*?\
        - The directory and file name cannot start or end with a space, but can contain spaces between them.
        - The HDFS full path contains a maximum of characters.
7. Click **OK**.

   You can view the file upload progress on the **File Operation Records** tab page. MRS processes the data import operation as a DistCp job. You can also check whether the DistCp job is successfully executed on the **Jobs** tab page.

Exporting Data

After the data analysis and computing are completed, you can store the data in HDFS or export them to OBS.

You can perform the following steps to export files and directories:

1. Log in to the MRS console.
2. Choose **Clusters > Active Clusters** and click the name of the cluster to be queried to enter the page displaying the cluster's basic information.
3. Click the **Files** tab, and the file management page is displayed.
4. Select **HDFS File List**.
5. Go to the data storage directory, for example, **bd_app1**.
6. Click **Export Data** and configure the OBS and HDFS paths. When configuring the OBS or HDFS path, click **Browse**, select a file directory, and click **Yes**.
    - OBS path
        - The path must start with **obs://**.
        - The directory and file name can contain letters, digits, hyphens (-), and underscores (_), but cannot contain the following special characters: ;|&>,<'$*?\
        - The directory and file name cannot start or end with a space, but can contain spaces between them.
        - The OBS full path contains a maximum of characters.
    - HDFS path
        - The path starts with **/user** by default.

- The directory and file name can contain letters, digits, hyphens (-), and underscores (_), but cannot contain the following special characters: ;|&>,<'$*?\
- The directory and file name cannot start or end with a space, but can contain spaces between them.
- The HDFS full path contains a maximum of characters.

NOTE

When a folder is exported to OBS, a label file named **folder name_$folder$** is added to the OBS path. Ensure that the exported folder is not empty. If the exported folder is empty, OBS cannot display the folder and only generates a file named **folder name_ $folder$**.

7. Click **OK**.

   You can view the file upload progress on the **File Operation Records** tab page. MRS processes the data export operation as a DistCp job. You can also check whether the DistCp job is successfully executed on the **Jobs** tab page.

Viewing Operation Logs

When importing and exporting data on the MRS management console, you can choose **Files > File Operation Records** to view the data import and export progress.

Table 4-16 describes the parameters of the file operation record.

Table 4-16 File operation record parameters

Parameter	Description
Submitted	Start time of data import or export.
Source Path	Source path of data. ● OBS path during data import. ● HDFS path during data export.
Target Path	Target path of data. ● HDFS path during data import. ● OBS path during data import.
Status	Status during data import or export. ● Submitted ● Accepted ● Running ● Completed ● Terminated ● Abnormal

Parameter	Description
Duration (min)	Time of data import or export. The unit is minute.
Result	Result of data import or export. ● Successful ● Failed ● Killed ● Undefined
Operation	View Log: allows you to view file operation logs.

4.3.2 Changing the Subnet of a Cluster

If the current subnet does not have sufficient IP addresses, you can change to another subnet in the same VPC of the current cluster to obtain more available subnet IP addresses. Changing a subnet does not affect the IP addresses or subnets of existing nodes.

For details about how to configure network ACL outbound rules, see **How Do I Configure a Network ACL Outbound Rule?**

Changing a Subnet When No Network ACL Is Associated

Step 1 Log in to the MRS console.

Step 2 Click the target cluster name to go to its details page.

Step 3 Click **Change Subnet** on the right of **Default Subnet**.

Step 4 Select the target subnet and click **OK**.

If no subnet is available, click **Create Subnet** to create a subnet first.

----End

Changing a Subnet When a Network ACL Is Associated

Step 1 Log in to the MRS console and click the target cluster to go to its details page.

Step 2 In the **Basic Information** area, view **VPC**.

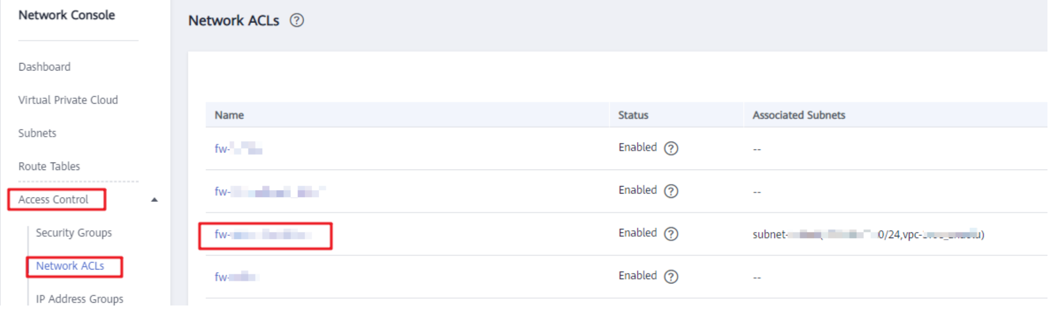
Step 3 Log in to the VPC console. In the navigation pane on the left, choose **Virtual Private Cloud** and obtain the IPv4 CIDR block corresponding to the VPC obtained in **Step 2**.

Step 4 Choose **Access Control** > **Network ACLs** and click the name of the network ACL that is associated with the default and new subnets.

NOTE

If both the default and new subnets are associated with a network ACL, add inbound rules to the network ACL by referring to **Step 5** to **Step 7**.

Figure 4-2 Network ACLs



Name	Status	Associated Subnets
fw-[redacted]	Enabled	--
fw-[redacted]	Enabled	--
fw-[redacted]	Enabled	subnet-[redacted]0/24,vpc-[redacted])
fw[redacted]	Enabled	--

Step 5 On the **Inbound Rules** page, choose **More** > **Insert Rule Above** in the **Operation** column.

Step 6 Add a network ACL rule. Set **Action** to **Allow**, **Source** to the VPC IPv4 CIDR block obtained in **Step 3**, and retain the default values for other parameters.

Step 7 Click **OK**.

NOTE

If you do not want to allow access from all IPv4 CIDR blocks of the VPC, add the IPv4 CIDR blocks of the default and new subnets by performing **Step 8** to **Step 12**. If the rules for VPC IPv4 CIDR blocks have been added, skip **Step 8** to **Step 12**.

Step 8 Log in to the MRS console.

Step 9 Click the target cluster to go to its details page.

Step 10 Click **Change Subnet** on the right of **Default Subnet**.

Step 11 Obtain the IPv4 CIDR blocks of the default and new subnets.

NOTICE

In this case, you do not need to click **OK** displayed in the **Change Subnet** dialog box. Otherwise, the default subnet will be updated to the new subnet, thereby making it difficult to query the IPv4 CIDR block of the default subnet. Exercise caution when performing this operation.

Step 12 Add the IPv4 CIDR blocks of the default and target subnets to the inbound rules of the network ACL bound to the two subnets by referring to **Step 4** to **Step 7**.

Step 13 Log in to the MRS console.

Step 14 Click the target cluster to go to its details page.

Step 15 Click **Change Subnet** on the right of **Default Subnet**.

Step 16 Select the target subnet and click **OK**.

----End

How Do I Configure a Network ACL Outbound Rule?

- Method 1

Allow all outbound traffic. This method ensures that clusters can be created and used properly.

- Method 2

  Allow the mandatory outbound rules that can ensure the successful creation of clusters. You are not advised to use this method because created clusters may not run properly due to absent outbound rules. If the preceding problem occurs, contact O&M personnel.

  Similar to the example provided in method 1, set **Action** to **Allow** and add the outbound rules whose destinations are the address with **Secure Communications** enabled, NTP server address, OBS server address, OpenStack address, and DNS server address, respectively.

4.3.3 Configuring Message Notification

MRS uses SMN to offer a publish/subscribe model to achieve one-to-multiple message subscriptions and notifications in a variety of message types (SMSs and emails).

Scenario

On the MRS management console, you can enable or disable the notification service on the **Alarms** page. The functions in the following scenarios can be implemented only after the required cluster function is enabled:

- After a user subscribes to the notification service, the MRS management plane notifies the user of success or failure of manual cluster scale-out and scale-in, cluster termination, and auto scaling by emails or SMS messages.
- The management plane checks the alarms about the MRS cluster and sends a notification to the tenant if the alarms are critical.
- If either of the operations such as deletion, shutdown, specifications modification, restart, and OS update is performed on an ECS in a cluster, the MRS cluster works abnormally. The management plane notifies a user when detecting that the VM of the user is in either of the preceding operations.

Creating a Topic

A topic is a specified event for message publication and notification subscription. It serves as a message sending channel, where publishers and subscribers can interact with each other.

1. Log in to the management console.
2. Click **Service List**. Under **Management & Governance**, click **Simple Message Notification**.

   The **SMN** page is displayed.
3. In the navigation pane, choose **Topic Management** > **Topics**.

   The **Topics** page is displayed.
4. Click **Create Topic**.

   The **Create Topic** dialog box is displayed.
5. In **Topic Name**, enter a topic name. In **Display Name**, enter a display name.

6. Select an existing project from the **Enterprise Project** drop-down list, or click **Create Enterprise Project** to create an enterprise project on the **Enterprise Project Management** page and then select it.
7. Set tag keys and tag values. Tags consist of keys and values. They identify cloud resources so that you can easily categorize and search for your resources.

Adding Subscriptions to a Topic

To deliver messages published to a topic to subscribers, you must add subscription endpoints to the topic. SMN automatically sends a confirmation message to the subscription endpoint. The confirmation message is valid only within 48 hours. The subscribers must confirm the subscription within 48 hours so that they can receive notification messages. Otherwise, the confirmation message becomes invalid, and you need to send it again.

1. Log in to the management console.
2. Under **Management & Governance**, click **Simple Message Notification**.

   The **SMN** page is displayed.
3. In the navigation pane, choose **Topic Management** > **Topics**.

   The **Topics** page is displayed.
4. Locate the topic to which you want to add a subscription, click **More** in the **Operation** column, and select **Add Subscription**.

   The **Add Subscription** box is displayed.

   Protocol can be set to **SMS**, FunctionGraph (function), **HTTP**, **HTTPS**, and **Email**.

   Endpoint indicates the address of the subscription endpoint. SMS and email, endpoints can be entered in batches. When adding endpoints in batches, each endpoint address occupies a line. You can enter a maximum of 10 endpoints.
5. Click **OK**.

The subscription you added is displayed in the subscription list.

Sending Notifications to Subscribers

1. Log in to the MRS console.
2. Choose **Clusters > Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.
3. Click **Alarms**.
4. Choose **Notification Rules** > **Add Notification Rule**. The **Add Notification Rule** page is displayed.
5. Set the notification rule parameters.

Table 4-17 Parameters of a notification rule

Parameter	Description
Rule Name	User-defined notification rule name. Only digits, letters, hyphens (-), and underscores (_) are allowed.
Message Notification	● If you enable this function, the system sends notifications to subscribers based on the notification rule. ● If you disable this function, the rule does not take effect, that is, notifications are not sent to subscribers.
Topic Name	Select an existing topic or click **Create Topic** to create a topic.
Notification Type	Select the type of the notification to be subscribed to. ● Alarm
Subscription Items	Select the items to be subscribed to. You can select all or some items as required. Subscription rules in MRS 3.*x* or later: Alarm severity: critical, major, and minor Subscription rules in versions earlier than MRS 3.x: ● Critical ● Major ● Minor ● Suggestion

6. Click **OK**.

4.3.4 Checking Health Status

4.3.4.1 Before You Start

This section describes how to manage health checks on the MRS console.

Health check management operations on the MRS console apply only to clusters of **MRS 1.9.2**.

Health check management on Manager applies to all versions. For MRS 3.x and later versions, see **Viewing a Health Check Task**. For versions earlier than MRS 3.x, see **Performing a Health Check**.

4.3.4.2 Performing a Health Check

Scenario

To ensure that cluster parameters, configurations, and monitoring are correct and that the cluster can run stably for a long time, you can perform a health check during routine maintenance.

NOTE

A system health check includes MRS Manager, service-level, and host-level health checks:

- MRS Manager health checks focus on whether the unified management platform can provide management functions.
- Service-level health checks focus on whether components can provide services properly.
- Host-level health checks focus on whether host indicators are normal.

The system health check includes three types of check items: health status, related alarms, and customized monitoring indicators for each check object. The health check results are not always the same as the **Health Status** on the portal.

Procedure

- Manually perform the health check for all services.

  On the MRS details page, choose **Management Operations** > **Start Cluster Health Check**.

  NOTE

  For the operations on MRS Manager, see **Performing a Health Check**; for the operations on FusionInsight Manager of MRS 3.*x* or later, see **Overview**.

  - The cluster health check includes Manager, service, and host status checks.
  - To perform cluster health checks, you can also choose **System** > **Check Health Status** > **Start Cluster Health Check** on MRS Manager.
  - To export the health check result, click **Export Report** in the upper left corner.
- Manually perform the health check for a service.
  1. On the MRS cluster details page, click **Components**.
  2. Select the target service from the service list.
  3. Choose **More** > **Start Service Health Check** to start the health check for the service.
- Manually perform the health check for a host.
  1. On the MRS details page, click **Nodes**.
  2. Expand the node group information and select the check box of the host to be checked.
  3. Choose **Node** > **Start Host Health Check** to start the health check for the host.

4.3.4.3 Viewing and Exporting a Health Check Report

Scenario

You can view the health check result on MRS and export it for further analysis.

NOTE

A system health check includes MRS Manager, service-level, and host-level health checks:

- MRS Manager health checks focus on whether the unified management platform can provide management functions.
- Service-level health checks focus on whether components can provide services properly.
- Host-level health checks focus on whether host indicators are normal.

The system health check includes three types of check items: health status, related alarms, and customized monitoring indicators for each check object. The health check results are not always the same as the **Health Status** on the portal.

Prerequisites

You have performed a health check.

Procedure

Step 1 On the MRS details page, choose **Management Operations** > **View Cluster Health Check Report**.

Step 2 Click **Export Report** on the health check report pane to export the report and view detailed information about check items.

----End

4.3.5 Remote O&M

4.3.5.1 Authorizing O&M

If you need technical support personnel to help you with troubleshooting, you can use the O&M authorization function to authorize technical support personnel to access your local host for fault location.

Procedure

Step 1 Log in to the MRS management console.

Step 2 In the navigation tree of the MRS management console, choose **Clusters** > **Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.

Step 3 In the upper right corner of the page, click **O&M**, choose **Authorize O&M**, and select the deadline for the support personnel to access the local host. Before the deadline, the support personnel have the temporary permission to access the local host.

Step 4 After the fault is rectified, click **O&M** in the upper right corner of the page and select **Cancel Authorization** to cancel the access permission for the support personnel.

----End

4.3.5.2 Sharing Logs

If you need technical support personnel to help you with troubleshooting, you can use the log sharing function to provide logs in a specific time to technical support personnel for fault location.

Procedure

Step 1 Log in to the MRS management console.

Step 2 In the navigation tree of the MRS management console, choose **Clusters** > **Active Clusters**, select a cluster, and click its name to switch to the cluster details page.

Step 3 In the upper right corner of the displayed page, choose **O&M** > **Share Log** to open the **Share Log** dialog box.

Step 4 Select the start time and end time in **Time Range**.

NOTE

- Select **Time Range** based on the suggestions of support personnel.
- **End Date** must be later than **Start Date**. Otherwise, logs cannot be filtered by time.

----End

4.3.6 Viewing MRS Operation Logs

You can view operation logs of clusters and jobs on the **Operation Logs** page. Log information is typically used for quickly locating faults in case of cluster exceptions, helping users resolve problems.

Operation Type

Currently, the following operation logs are provided by MRS. You can filter the logs in the search box.

- Cluster operations
  - Creating, deleting, scaling out, and scaling in a cluster
  - Creating and deleting a directory, deleting a file
- Job operations: Creating, stopping, and deleting a job
- Data operations: IAM user tasks, adding user, and adding user group

Log Fields

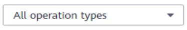
Logs are listed in chronological order by default in the log list, with the most recent logs displayed at the top.

Table 4-18 describes various fields in a log.

Table 4-18 Log description

Parameter	Description
Operation Type	Various types of operations, including: ● Cluster operations ● Job operations ● Data operations
Operation IP	IP address where an operation is performed. **NOTE** If an MRS cluster fails to be deployed, the cluster is automatically deleted, and the operation logs of the automatically deleted cluster do not contain the **Operation IP** of the user.
Operation	Operation details. The value can contain a maximum of 2048 characters.
Time	Operation time. For a deleted cluster, only logs generated within the last six months are displayed. To view logs generated six months ago, contact technical support.

Table 4-19 Icon description

Icon	Description
 All operation types ▾	Select an operation type from the drop-down list box to filter logs. ● **All Operation Types**: Filter all logs. ● **Cluster**: Filter logs for **Cluster**. ● **Job**: Filter logs for **Job**. ● **Data**: Filter logs for **Data**.
📅	Filter logs by time. 1. Click the input box. 2. Specify the date and time. 3. Click **OK**. The left-side input box indicates the start time and the right-side one indicates the end time. The start time must be earlier than or equal to the end time. Otherwise, logs cannot be filtered.
🔍	Enter a keyword of the **Operation Details** in the search box and click 🔍 to search for logs.
↻	Click ↻ to manually refresh the log list.

4.3.7 Terminating a Cluster

You can terminate an MRS cluster after job execution is complete.

Background

You can manually terminate a cluster after data analysis is complete or when the cluster encounters an exception. A cluster failed to be deployed will be automatically terminated.

Procedure

Step 1 Log in to the MRS management console.

Step 2 In the navigation pane on the left, choose **Clusters** > **Active Clusters**.

Step 3 In the cluster list, locate the row containing the cluster to be terminated, and click **Terminate** in the **Operation** column.

The cluster status changes from **Running** to **Terminating**, and finally to **Terminated**. You can view the terminated cluster in **Cluster History**.

----End

4.4 Managing Nodes

4.4.1 Manually Scaling Out a Cluster

The storage and computing capabilities of MRS can be improved by simply adding Core nodes or Task nodes instead of modifying system architecture, reducing O&M costs. Core nodes can process and store data. You can add Core nodes to expand the node quantities and handle peak loads. Task nodes are used for computing and do not store persistent data.

Background

The MRS cluster supports a maximum of 500 Core and Task nodes. If more than 500 Core/Task nodes are required, contact technical support engineers or invoke a background interface to modify the database.

Core nodes and Task nodes can be added, excluding the Master node. Here, the maximum number of Core/Task nodes to be added is 500 minus the number of Core/Task nodes. For example, the current number of Core nodes is 3, the number of Core nodes to be added must be less than or equal to 497. If the cluster scale-out fails, you can add node to the cluster again.

If no node is added during cluster creation, you can specify the number of nodes to be added during scale-out. However, you cannot specify the nodes to be added.

The operations for scaling out a cluster vary depending on the selected version.

Procedure

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters** > **Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.

Step 3 Click the **Nodes** tab. In the **Operation** column of the node group, click **Scale Out**. The **Scale Out** page is displayed.

The scale-out operation can only be performed on the running clusters.

Step 4 Set **Scaled Out Nodes**, **Enable Component**, and **Run Bootstrap Action**, and click **OK**

NOTE

- If the Task node group does not exist in the cluster, configure the Task node by referring to **Adding a Task Node**.
- If a bootstrap action is added during cluster creation, the **Run Bootstrap Action** parameter is valid. If this function is enabled, the bootstrap actions added during cluster creation will be run on all the scaled out nodes.
- If the **New Specifications** parameter is available, the specifications that are the same as those of the original nodes have been sold out or discontinued. Nodes with new specifications will be added.
- Before scaling out the cluster, check whether its security group configuration is correct. Ensure that an inbound security group rule contains a rule in which **Protocol & Port** is set to **All**, and **Source** is set to a trusted accessible IP address range.

Step 5 In the **Scale Out Node** dialog box, click **OK**.

Step 6 A dialog box is displayed, indicating that the scale-out task is submitted successfully.

The following parameters explain the cluster scale-out process:

- Expanding: If a cluster is being expanded, its status is **Scaling out**. The submitted jobs will be executed and you can submit new jobs. You are not allowed to continue to scale out, or delete the cluster. You are advised not to restart the cluster or modify the cluster configuration.
- Expansion succeeded: If a cluster is expanded successfully, its status is **Running**.
- Failed scale-out: The cluster status is **Running** when the cluster scale-out failed. You can execute jobs and scale out the cluster again.

After the cluster is scaled out, you can view the node information of the cluster on the **Nodes** page.

----End

Adding a Task Node

To add a task node to a custom cluster, perform the following steps:

1. On the cluster details page, click the **Nodes** tab and click **Add Node Group**. The **Add Node Group** page is displayed.
2. Select **NM** for **Deploy Roles** and set other parameters as required.

To add a task node to a non-custom cluster, perform the following steps:

1. On the cluster details page, click the **Nodes** tab and click **Configure Task Node**. The **Configure Task Node** page is displayed.
2. On the **Configure Task Node** page, set **Node Type**, **Instance Specifications**, **Nodes**, **System Disk**. In addition, if **Add Data Disk** is enabled, configure the storage type, size, and number of data disks.
3. Click **OK**.

Adding a Node Group

NOTE

Used to add node groups and applies to customized clusters of MRS 3.*x*.

1. On the cluster details page, click the **Nodes** tab and click **Add Node Group**. The **Add Node Group** page is displayed.
2. Set the parameters as needed.

Table 4-20 Parameters for adding a node group

Parameter	Description
Instance Specifications	Select the flavor type of the hosts in the node group.
Nodes	Set the number of nodes in the node group.
System Disk	Set the specifications and capacity of the system disk on the new node.
Data Disk (GB)	Set the specifications, capacity, and number of data disks of the new node.
Deploy Roles	Deploy the instances of each node in the new node group. The setting can be manually adjusted.

3. Click **OK**.

4.4.2 Manually Scaling In a Cluster

You can reduce the number of core or task nodes to scale in a cluster based on service requirements so that MRS delivers better storage and computing capabilities at lower O&M costs.

The scale-in operation is not allowed for a cluster that is performing active/standby synchronization.

Background

A cluster can have three types of nodes, master, core, and task nodes. Currently, only core and task nodes can be removed. To scale in a cluster, you only need to adjust the number of nodes on the MRS console. MRS then automatically selects the nodes to be removed.

The policies for MRS to automatically select nodes are as follows:

- MRS does not select the nodes with basic components installed, such as ZooKeeper, DBService, KrbServer, and LdapServer, because these basic components are the basis for the cluster to run.
- Core nodes store cluster service data. When scaling in a cluster, ensure that all data on the core nodes to be removed has been migrated to other nodes. You can perform follow-up scale-in operations only after all component services are decommissioned, for example, removing nodes from Manager and deleting ECSs. When selecting core nodes, MRS preferentially selects the nodes with a small amount of data and healthy instances to be decommissioned to prevent decommissioning failures. For example, if DataNodes are installed on core nodes in an analysis cluster, MRS preferentially selects the nodes with small data volume and good health status during scale-in.

  When core nodes are removed, their data is migrated to other nodes. If the user business has cached the data storage path, the client will automatically update the path, which may increase the service processing latency temporarily. Cluster scale-in may slow the response of the first access to some HBase on HDFS data. You can restart HBase or disable or enable related tables to resolve this issue.
- Task nodes are computing nodes and do not store cluster data. Data migration is not involved in removing task nodes. Therefore, when selecting task nodes, MRS preferentially selects nodes whose health status is faulty, unknown, or subhealthy. On the **Components** tab of the MRS console, click a service and then the **Instances** tab to view the health status of the node instances.

Scale-In Verification Policy

To prevent component decommissioning failures, components provide different decommissioning constraints. Scale-in is allowed only when the constraints of all installed components are met. **Table 4-21** describes the scale-in verification policies.

Table 4-21 Decommissioning constraints

Component	Constraint
HDFS/DataNode	The number of available nodes after the scale-in is greater than or equal to the number of HDFS copies and the total HDFS data volume does not exceed 80% of the total HDFS cluster capacity. This ensures that the remaining space is sufficient for storing existing data after the scale-in and reserves some space for future use. **NOTE** To ensure data reliability, one backup is automatically generated for each file saved in HDFS, that is, two copies are generated in total.

Component	Constraint
HBase/ RegionServer	The total available memory of RegionServers on all nodes except the nodes to be removed is greater than 1.2 times of the memory which is currently used by RegionServers on these nodes. This ensures that the node to which the region on a decommissioned node is migrated has sufficient memory to bear the region of the decommissioned node.
Storm/ Supervisor	After the scale-in, ensure that the number of slots in the cluster is sufficient for running the submitted tasks. This prevents no sufficient resources being available for running the stream processing tasks after the scale-in.
Flume/ FlumeServer	If FlumeServer is installed on a node and Flume tasks have been configured for the node, the node cannot be deleted. This prevents the deployed service program from being deleted by mistake.

Scaling In a Cluster by Specifying the Node Quantity

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters** > **Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.

Step 3 Click the **Nodes** tab. In the **Operation** column of the node group, click **Scale In** to go to the **Scale In** page.

This operation can be performed only when the cluster and all nodes in it are running.

Step 4 Set **Scale-In Nodes** and click **OK**.

NOTE

- Before scaling in the cluster, check whether its security group configuration is correct. Ensure that an inbound security group rule contains a rule in which **Protocol & Port** is set to **All**, and **Source** is set to a trusted accessible IP address range.
- If damaged data blocks exist in HDFS, the cluster may fail to be scaled in. Contact technical support.

Step 5 A dialog box displayed in the upper right corner of the page indicates that the task of removing the node is submitted successfully.

The cluster scale-in process is explained as follows:

- During scale-in: The cluster status is **Scaling In**. The submitted jobs will be executed, and you can submit new jobs. You are not allowed to continue to scale in or terminate the cluster. You are advised not to restart the cluster or modify the cluster configuration.
- Successful scale-in: The cluster status is **Running**.
- Failed scale-in: The cluster status is **Running**. You can execute jobs or scale-in the cluster again.

After the cluster is scaled in, you can view the node information of the cluster on the **Nodes** page.

----End

4.4.3 Managing a Host (Node)

Scenario

To check an abnormal or faulty host (node), you need to stop all host roles on MRS. To recover host services after the host fault is rectified, restart all roles.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS details page, click **Nodes**.

Step 2 Unfold the node group information and select the check box of the target node.

Step 3 Choose **Node Operation** > **Start All Roles** or **Stop All Roles** to perform the required operation.

----End

4.4.4 Isolating a Host

Scenario

If a host is found to be abnormal or faulty, affecting cluster performance or preventing services from being provided, you can temporarily exclude that host from the available nodes in the cluster. In this way, the client can access other available nodes. In scenarios where patches are to be installed in a cluster, you can also exclude a specified node from patch installation.

You can isolate a host manually on MRS based on the actual service requirements or O&M plan. Only non-management nodes can be isolated.

Impact on the System

- After a host is isolated, all role instances on the host will be stopped. You cannot start, stop, or configure the host and any instances on the host.
- After a host is isolated, statistics of the monitoring status and indicator data of the host hardware and instances cannot be collected or displayed.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS details page, click **Nodes**.

Step 2 Unfold the node group information and select the check box of the target host.

Step 3 Choose **Node Operation** > **Isolate Host**.

Step 4 Confirm the information about the host to be isolated and click **OK**.

When **Operation successful** is displayed, click **Finish**. The host is isolated successfully, and the value of **Operating Status** becomes **Isolated**.

NOTE

For isolated hosts, you can cancel the isolation and add them to the cluster again. For details, see **Canceling Host Isolation**.

----End

4.4.5 Canceling Host Isolation

Scenario

After the exception or fault of a host is handled, you must cancel the isolation of the host for proper usage.

You can cancel the isolation of a host on MRS.

Prerequisites

- The host is in the **Isolated** state.
- The exception or fault of the host has been rectified.
- You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS details page, click **Nodes**.

Step 2 Unfold the node group information and select the check box of the target host that you want to cancel its isolation.

Step 3 Choose **Node Operation** > **Cancel Host Isolation**.

Step 4 Confirm the information about the host for which the isolation is to be cancelled and click **OK**.

When **Operation successful** is displayed, click **Finish**. The host is de-isolated successfully, and the value of **Operating Status** becomes **Normal**.

Step 5 Select the host that has been de-isolated and choose **Node Operation** > **Start All Roles**.

----End

4.5 Job Management

4.5.1 Introduction to MRS Jobs

An MRS job is the program execution platform of MRS. It is used to process and analyze user data. After a job is created, all job information is displayed on the **Jobs** tab page. You can view a list of all jobs and create and manage jobs. If the **Jobs** tab is not displayed on the cluster details page, submit a job in the background.

Data sources processed by MRS are from OBS or HDFS. OBS is an object-based storage service that provides you with massive, secure, reliable, and cost-effective data storage capabilities. MRS can process data in OBS directly. You can view, manage, and use data by using the web page of the management control platform or OBS client. In addition, you can use REST APIs independently or integrate APIs to service applications to manage and access data.

Before creating jobs, upload the local data to OBS for MRS to compute and analyze. MRS allows exporting data from OBS to HDFS for computing and analyzing. After the analyzing and computing are complete, you can store the data in HDFS or export them to OBS. HDFS and OBS can also store the compressed data in the format of **bz2** or **gz**.

Category

An MRS cluster allows creating and managing the following jobs: If a cluster in the **Running** state fails to create a job, check the health status of related components on the cluster management page. For details, see **Viewing and Customizing Cluster Monitoring Metrics**.

- MapReduce: provides the capability of processing massive data quickly and in parallel. It is a distributed data processing mode and execution environment. MRS supports the submission of MapReduce JAR programs.
- Spark: a distributed in-memory computing framework. MRS supports SparkSubmit, Spark Script, and Spark SQL jobs.
  - SparkSubmit: You can submit the Spark JAR and Spark Python programs, execute the Spark Application, and compute and process user data.
  - SparkScript: You can submit the SparkScript scripts and batch execute Spark SQL statements.
  - Spark SQL: You can use Spark SQL statements (similar to SQL statements) to query and analyze user data in real time.
- Hive: an open-source data warehouse based on Hadoop. MRS allows you to submit HiveScript scripts and execute Hive SQL statements.
- Flink: provides a distributed big data processing engine that can perform stateful computations over both finite and infinite data streams.

Job List

Tasks are listed in chronological order by default in the task list, with the most recent jobs displayed at the top. **Table 4-22** describes the parameters in the job list.

Table 4-22 Job list parameters

Parameter	Description
Name/ID	Job name, which is set when a job is created. ID is the unique identifier of a job. After a job is added, the system automatically assigns a value to ID.
Username	Name of the user who submits a job.
Type	The following data types are supported: ● DistCp: importing and exporting data ● MapReduce ● Spark ● SparkSubmit ● SparkScript ● Spark SQL ● Hive SQL ● HiveScript ● Flink **NOTE** ● After importing and exporting files on the **Files** tab page, you can view the DistCp job on the **Jobs** tab page. ● Spark, Hive, and Flink jobs can be added only when the Spark, Hive, and Flink components are selected during cluster creation and the cluster is running.
Status	Job status. ● Submitted ● Accepted ● Running ● Completed ● Terminated ● Abnormal

Parameter	Description
Result	Execution result of a job. ● Undefined: indicates that the job is being executed. ● **Successful**: indicates that the job has been successfully executed. ● **Killed**: indicates that the job is manually terminated during execution. ● **Failed**: indicates that the job fails to be executed. **NOTE** Once a job has succeeded or failed, you cannot execute it again. However, you can add a job, and set job parameters to submit a job again.
Submitted	Time when a job is submitted.
Ended	Time when a job is completed or manually stopped.
Operation	● Viewing Log: Click **View Log** to view the real-time logs of running jobs. For details, see **Viewing Job Configuration and Logs**. ● View Details: Click **View Details** to view the detailed configuration information about jobs. For details, see **Viewing Job Configuration and Logs**. ● More – Stop: You can click **Stop** to stop a running job. For details, see **Stopping a Job**. – Delete: Click **Delete** to delete a job. For details, see **Deleting a Job**. – View Result: Click **View Result** to view the execution results of SparkSQL and SparkScript jobs whose status is **Completed** and result is **Successful**. **NOTE** ● You cannot stop Spark SQL jobs. ● A deleted job cannot be restored. Therefore, exercise caution when deleting a job. ● If you choose to save job logs to OBS or HDFS, the system compresses and saves the logs to the corresponding path after the job execution is completed. Therefore, after a job execution of this type is completed, the job status is still **Running**. After the log is successfully stored, the job status changes to **Completed**. The log storage duration depends on the log size and takes several minutes.

Table 4-23 Icon description

Icon	Description
📅	Select a time range for job submission to filter jobs submitted in the time range.
All statuses ▾	Select a certain job execution result from the drop-down list to display jobs of the status. ● All statuses: Filter all jobs. ● Successful: Filter jobs that are successfully executed. ● Undefined: Filter jobs that are being executed. ● Killed: Filter jobs that are manually stopped. ● Failed: Filter jobs that fail to be executed.
All types ▾	Select a certain job type from the drop-down list to display jobs of the type. ● All types ● MapReduce ● HiveScript ● Distcp ● SparkScript ● Spark SQL ● Hive SQL ● SparkSubmit ● Flink
Name ▾	In the search box, search for a job by setting the corresponding search condition and click 🔍. ● Job name. ● Job ID. ● Username. ● Queue name.
↻	Click ↻ to manually refresh the job list.

Job Execution Permission Description

For a security cluster with Kerberos authentication enabled, a user needs to synchronize an IAM user before submitting a job on the MRS web UI. After the synchronization is completed, the MRS system generates a user with the same IAM username. Whether a user has the permission to submit jobs depends on the IAM policy bound to the user during IAM synchronization. For details about the job submission policy, see **Table 2-3** in **Synchronizing IAM Users to MRS**.

When a user submits a job that involves the resource usage of a specific component, such as accessing HDFS directories and Hive tables, user **admin**

(Manager administrator) must grant the relevant permission to the user. Detailed operations are as follows:

Step 1 Log in to Manager as user **admin**.

Step 2 Add the role of the component whose permission is required by the user. For details, see **Creating a Role**.

Step 3 Change the user group to which the user who submits the job belongs and add the new component role to the user group. For details, see **Related Tasks**.

NOTE

After the component role bound to the user group to which the user belongs is modified, it takes some time for the role permissions to take effect.

----End

4.5.2 Running a MapReduce Job

You can submit programs developed by yourself to MRS to execute them, and obtain the results. This section describes how to submit a MapReduce job on the MRS management console. MapReduce jobs are used to submit JAR programs to quickly process massive amounts of data in parallel and create a distributed data processing and execution environment.

If the job and file management functions are not supported on the cluster details page, submit the jobs in the background.

Prerequisites

You have uploaded the program packages and data files required for running jobs to OBS or HDFS.

Before you upload the program packages and data files to OBS, you need to create an OBS agency and bind it to the MRS cluster. For details, see **Configuring a Storage-Compute Decoupled Cluster (Agency)**.

Submitting a Job on the GUI

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters > Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.

Step 3 If Kerberos authentication is enabled for the cluster, perform the following steps. If Kerberos authentication is not enabled for the cluster, skip this step.

In the **Basic Information** area on the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users. For details, see **Synchronizing IAM Users to MRS**.

NOTE

- When the policy of the user group to which the IAM user belongs changes from MRS ReadOnlyAccess to MRS CommonOperations, MRS FullAccess, or MRS Administrator, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** (System Security Services Daemon) cache of cluster nodes needs time to be updated. Then, submit a job. Otherwise, the job may fail to be submitted.
- When the policy of the user group to which the IAM user belongs changes from MRS CommonOperations, MRS FullAccess, or MRS Administrator to MRS ReadOnlyAccess, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** cache of cluster nodes needs time to be updated.

Step 4 Click the **Jobs** tab.

Step 5 Click **Create**. The **Create Job** page is displayed.

NOTE

If the IAM username contains spaces (for example, **admin 01**), a job cannot be created.

Step 6 In **Type**, select **MapReduce**. Configure other job information.

- Configure MapReduce job information by referring to **Table 4-26** if the cluster version is earlier than MRS 2.0.1.

Table 4-24 Job configuration information

Parameter	Description
Name	Job name. It contains 1 to 64 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed. **NOTE** You are advised to set different names for different jobs.
Program Path	Path of the program package to be executed. The following requirements must be met: • Contains a maximum of 1,023 characters, excluding special characters such as ;|&><'$. The parameter value cannot be empty or full of spaces. • The path of the program to be executed can be stored in HDFS or OBS. The path varies depending on the file system. – OBS: The path must start with **obs://**. Example: **obs://wordcount/program/*xxx*.jar** – HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. • For SparkScript and HiveScript, the path must end with **.sql**. For MapReduce, the path must end with **.jar**. For Flink and SparkSubmit, the path must end with **.jar** or **.py**. The **.sql**, **.jar**, and **.py** are case-insensitive.

Parameter	Description
Parameters	(Optional) It is the key parameter for program execution. Multiple parameters are separated by space. Configuration method: *Program class name Data input path Data output path* ● Program class name: It is specified by a function in your program. MRS is responsible for transferring parameters only. ● Data input path: Click **HDFS** or **OBS** to select a path or manually enter a correct path. ● Data output path: Enter a directory that does not exist. The parameter contains a maximum of 150,000 characters. It cannot contain special characters ;|&><'$, but can be left blank. **CAUTION** If you enter a parameter with sensitive information (such as the login password), the parameter may be exposed in the job details display and log printing. Exercise caution when performing this operation.
Service Parameter	(Optional) It is used to modify service parameters for the job. The parameter modification applies only to the current job. To make the modification take effect permanently for the cluster, follow instructions in **Configuring Service Parameters**. To add multiple parameters, click ⊕ on the right. To delete a parameter, click **Delete** on the right. **Table 4-25** lists the common service configuration parameters.
Command Reference	Command submitted to the background for execution when a job is submitted.

Table 4-25 Service Parameter parameters

Parameter	Description	Example Value
fs.obs.access.key	Key ID for accessing OBS.	-
fs.obs.secret.key	Key corresponding to the key ID for accessing OBS.	-

Table 4-26 Job configuration information

Parameter	Description
Name	Job name. It contains 1 to 64 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed. **NOTE** You are advised to set different names for different jobs.
Program Path	Path of the program package to be executed. The following requirements must be met: ● Contains a maximum of 1,023 characters, excluding special characters such as ;|&><'$. The parameter value cannot be empty or full of spaces. ● The path of the program to be executed can be stored in HDFS or OBS. The path varies depending on the file system. – OBS: The path must start with **s3a://**. Example: **s3a://wordcount/program/xxx.jar** – HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. ● For SparkScript, the path must end with **.sql**. For MapReduce and Spark, the path must end with **.jar**. The **.sql** and **.jar** are case-insensitive.
Parameters	Key parameter for program execution. The parameter is specified by the function of the user's program. MRS is only responsible for loading the parameter. Multiple parameters are separated by space. Configuration method: *Package name*.*Class name* The parameter contains a maximum of 150,000 characters. It cannot contain special characters ;|&><'$, but can be left blank. **NOTE** When entering a parameter containing sensitive information (for example, login password), you can add an at sign (@) before the parameter name to encrypt the parameter value. This prevents the sensitive information from being persisted in plaintext. When you view job information on the MRS management console, the sensitive information is displayed as *. Example: **username=admin @password=admin_123**
Import From	Path for inputting data Data can be stored in HDFS or OBS. The path varies depending on the file system. ● OBS: The path must start with **s3a://**. ● HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.

Parameter	Description
Export To	Path for outputting data **NOTE** ● When setting this parameter, select **OBS** or **HDFS**. Select a file directory or manually enter a file directory, and click **OK**. ● If you add the **hadoop-mapreduce-examples-x.x.x.jar** sample program or a program similar to **hadoop-mapreduce-examples-x.x.x.jar**, enter a directory that does not exist. Data can be stored in HDFS or OBS. The path varies depending on the file system. ● OBS: The path must start with **s3a://**. (Supported only in MRS 1.8.10 and earlier versions) ● HDFS: The path must start with **/user**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.
Log Path	Path for storing job logs that record job running status. Data can be stored in HDFS or OBS. The path varies depending on the file system. ● OBS: The path must start with **s3a://**. ● HDFS: The path must start with **/user**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.

Step 7 Confirm job configuration information and click **OK**.

After the job is created, you can manage it.

----End

Submitting a Job in the Background

The default client installation path for MRS 3.x or later is /opt/Bigdata/client, and for versions earlier than MRS 3.x is /opt/client. Configure the path based on site requirements.

Step 1 Log in to a Master node. For details, see **Logging In to an ECS**.

Step 2 Run the following command to initialize environment variables:

source /opt/Bigdata/client/bigdata_env

Step 3 If the Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If the Kerberos authentication is disabled for the current cluster, skip this step.

kinit *MRS cluster user*

Example: **kinit admin**

Step 4 Run the following command to copy the program in the OBS file system to the Master node in the cluster:

hadoop fs -Dfs.obs.access.key=AK -Dfs.obs.secret.key=SK -copyToLocal source_path.jar target_path.jar

Example: **hadoop fs -Dfs.obs.access.key=XXXX -Dfs.obs.secret.key=XXXX -copyToLocal "obs://mrs-word/program/hadoop-mapreduce-examples-XXX.jar" "/home/omm/hadoop-mapreduce-examples-XXX.jar"**

You can log in to OBS Console using AK/SK. To obtain AK/SK information, click the username in the upper right corner of the management console and choose **My Credentials** > **Access Keys**.

Step 5 Run the following command to submit a wordcount job. If data needs to be read from OBS or outputted to OBS, the AK/SK parameters need to be added.

source /opt/Bigdata/client/bigdata_env;hadoop jar execute_jar wordcount input_path output_path

Example: **source /opt/Bigdata/client/bigdata_env;hadoop jar /home/omm/hadoop-mapreduce-examples-XXX.jar wordcount -Dfs.obs.access.key=XXXX -Dfs.obs.secret.key=XXXX "obs://mrs-word/input/*" "obs://mrs-word/output/"**

In the preceding command, **input_path** indicates a path for storing job input files on OBS. **output_path** indicates a path for storing job output files on OBS and needs to be set to a directory that does not exist

----End

4.5.3 Running a SparkSubmit Job

You can submit programs developed by yourself to MRS to execute them, and obtain the results. This section describes how to submit a Spark job on the MRS console.

Prerequisites

You have uploaded the program packages and data files required for running jobs to OBS or HDFS.

Submitting a Job on the GUI

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters > Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.

Step 3 If Kerberos authentication is enabled for the cluster, perform the following steps. If Kerberos authentication is not enabled for the cluster, skip this step.

In the **Basic Information** area on the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users. For details, see **Synchronizing IAM Users to MRS**.

NOTE

- When the policy of the user group to which the IAM user belongs changes from MRS ReadOnlyAccess to MRS CommonOperations, MRS FullAccess, or MRS Administrator, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** (System Security Services Daemon) cache of cluster nodes needs time to be updated. Then, submit a job. Otherwise, the job may fail to be submitted.
- When the policy of the user group to which the IAM user belongs changes from MRS CommonOperations, MRS FullAccess, or MRS Administrator to MRS ReadOnlyAccess, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** cache of cluster nodes needs time to be updated.

Step 4 Click the **Jobs** tab.

Step 5 Click **Create**. The **Create Job** page is displayed.

Step 6 Configure job information.

- Set **Type** to **SparkSubmit** if the cluster version is or later. Configure other parameters of the SparkSubmit job by referring to **Table 4-27**.
- Set **Type** to **Spark**. Configure other parameters of the Spark job by referring to **Table 4-30**.

Table 4-27 Job configuration information

Parameter	Description
Name	Job name. It contains 1 to 64 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed. **NOTE** You are advised to set different names for different jobs.
Program Path	Path of the program package to be executed. The following requirements must be met: ● Contains a maximum of 1,023 characters, excluding special characters such as ;|&><'$. The parameter value cannot be empty or full of spaces. ● The path of the program to be executed can be stored in HDFS or OBS. The path varies depending on the file system. – OBS: The path must start with **obs://**. Example: **obs://wordcount/program/*xxx*.jar** – HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. ● For SparkScript and HiveScript, the path must end with **.sql**. For MapReduce, the path must end with **.jar**. For Flink and SparkSubmit, the path must end with **.jar** or **.py**. The **.sql**, **.jar**, and **.py** are case-insensitive.
Program Parameter	(Optional) Used to configure optimization parameters such as threads, memory, and vCPUs for the job to optimize resource usage and improve job execution performance. **Table 4-28** describes the common parameters of a running program.

Parameter	Description
Parameters	(Optional) Key parameter for program execution. The parameter is specified by the function of the user's program. MRS is only responsible for loading the parameter. Multiple parameters are separated by space. The parameter contains a maximum of 150,000 characters. It cannot contain special characters ;|&><'$, but can be left blank. **CAUTION** If you enter a parameter with sensitive information (such as the login password), the parameter may be exposed in the job details display and log printing. Exercise caution when performing this operation.
Service Parameter	(Optional) It is used to modify service parameters for the job. The parameter modification applies only to the current job. To make the modification take effect permanently for the cluster, follow instructions in **Configuring Service Parameters**. To add multiple parameters, click ⊕ on the right. To delete a parameter, click **Delete** on the right. **Table 4-29** lists the common service configuration parameters. **NOTE** If you need to run a long-term job, such as SparkStreaming, and access OBS, you need to use **Service Parameter** to import the AK/SK for accessing OBS.
Command Reference	Command submitted to the background for execution when a job is submitted.

Table 4-28 Program parameters

Parameter	Description	Example Value
--conf	Add the task configuration items.	spark.executor.memory=2G
--driver-memory	Set the running memory of driver.	2G
--num-executors	Set the number of executors to be started.	5
--executor-cores	Set the number of executor cores.	2
--class	Set the main class of a task.	org.apache.spark.examples.SparkPi
--files	Upload files to a task. The files can be custom configuration files or some data files from OBS or HDFS.	-
--jars	Upload additional dependency packages of a task to add the external dependency packages to the task.	-

Parameter	Description	Example Value
--executor-memory	Set executor memory.	2G
--conf spark-yarn.maxAppAttempts	Control the number of AM retries.	If this parameter is set to **0**, retry is not allowed. If this parameter is set to **1**, one retry is allowed.

Table 4-29 Service Parameter parameters

Parameter	Description	Example Value
fs.obs.access.key	Key ID for accessing OBS.	-
fs.obs.secret.key	Key corresponding to the key ID for accessing OBS.	-

Table 4-30 Job configuration information

Parameter	Description
Name	Job name. It contains 1 to 64 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed. **NOTE** You are advised to set different names for different jobs.
Program Path	Path of the program package to be executed. The following requirements must be met: ● Contains a maximum of 1,023 characters, excluding special characters such as ;|&><'$. The parameter value cannot be empty or full of spaces. ● The path of the program to be executed can be stored in HDFS or OBS. The path varies depending on the file system. – OBS: The path must start with **s3a://**. Example: **s3a://wordcount/program/*xxx*.jar** – HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. ● For SparkScript, the path must end with **.sql**. For MapReduce and Spark, the path must end with **.jar**. The **.sql** and **.jar** are case-insensitive.

Parameter	Description
Parameters	Key parameter for program execution. The parameter is specified by the function of the user's program. MRS is only responsible for loading the parameter. Multiple parameters are separated by space. Configuration method: *Package name*.*Class name* The parameter contains a maximum of 150,000 characters. It cannot contain special characters ;|&><'$, but can be left blank. **NOTE** When entering a parameter containing sensitive information (for example, login password), you can add an at sign (@) before the parameter name to encrypt the parameter value. This prevents the sensitive information from being persisted in plaintext. When you view job information on the MRS console, the sensitive information is displayed as *****. Example: **username=admin @password=admin_123**
Import From	Path for inputting data Data can be stored in HDFS or OBS. The path varies depending on the file system. ● OBS: The path must start with **s3a://**. ● HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.
Export To	Path for outputting data **NOTE** ● When setting this parameter, select **OBS** or **HDFS**. Select a file directory or manually enter a file directory, and click **OK**. ● If you add the **hadoop-mapreduce-examples-x.x.x.jar** sample program or a program similar to **hadoop-mapreduce-examples-x.x.x.jar**, enter a directory that does not exist. Data can be stored in HDFS or OBS. The path varies depending on the file system. ● OBS: The path must start with **s3a://**. ● HDFS: The path must start with **/user**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.

Parameter	Description
Log Path	Path for storing job logs that record job running status. Data can be stored in HDFS or OBS. The path varies depending on the file system. ● OBS: The path must start with **s3a://**. ● HDFS: The path must start with **/user**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.

Step 7 Confirm job configuration information and click **OK**.

After the job is created, you can manage it.

----End

Submitting a Job in the Background

The default client installation path for MRS 3.x or later is /opt/Bigdata/client, and for versions earlier than MRS 3.x is /opt/client. Configure the path based on site requirements.

Step 1 Create a user for submitting jobs. For details, see **Creating a User**.

In this example, a machine-machine user used in the user development scenario has been created, and user groups (**hadoop** and **supergroup**), the primary group (**supergroup**), and role permissions (**System_administrator** and **default**) have been correctly assigned to the user.

Step 2 Download the authentication credential.

- For clusters of MRS 3.*x* or later, log in to FusionInsight Manager and choose **System** > **Permission** > **User**. In the **Operation** column of the newly created user, choose **More** > **Download Authentication Credential**.
- For clusters whose version is earlier than MRS 3.*x*, log in to MRS Manager and choose **System** > **Manage User**. In the **Operation** column of the newly created user, choose **More** > **Download Authentication Credential**.

Step 3 Upload JAR files related to the job to the cluster. In this example, the sample JAR file built in Spark is used. It is stored in **$SPARK_HOME/examples/jars**.

Step 4 Upload the authentication credential of the user created in **Step 2** to the **/opt** directory of the cluster and run the following command to decompress the credential:

tar –xvf MRSTest _*xxxxxx*_keytab.tar

You will obtain two files: **user.keytab** and **krb5.conf**.

Step 5 Before performing operations on the cluster, run the following commands:

source /opt/Bigdata/client/bigdata_env

cd $SPARK_HOME

Step 6 Run the following command to submit the Spark job:

./bin/spark-submit --master yarn --deploy-mode client --conf spark.yarn.principal=MRSTest --conf spark.yarn.keytab=/opt/user.keytab --class org.apache.spark.examples.SparkPi examples/jars/spark-examples_2.11-2.3.2-mrs-2.0.jar 10

Parameter description:

1. Computing capability of Yarn, which specifies that the job is submitted in client mode.
2. Configuration item of the Spark job. The authentication file and username are transferred here.
3. **spark.yarn.principal**: user created in step 1
4. **spark.yarn.keytab**: keytab file used for authentication
5. *xx***.jar**: JAR file used by the job

----End

4.5.4 Running a HiveSQL Job

You can submit programs developed by yourself to MRS to execute them, and obtain the results. This section describes how to submit a HiveSQL job on the MRS management console. HiveSQL jobs are used to submit SQL statements and script files for data query and analysis. Both SQL statements and scripts are supported. If SQL statements contain sensitive information, use Script to submit them.

Prerequisites

You have uploaded the program packages and data files required for running jobs to OBS or HDFS.

Submitting a Job on the GUI

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters > Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.

Step 3 If Kerberos authentication is enabled for the cluster, perform the following steps. If Kerberos authentication is not enabled for the cluster, skip this step.

In the **Basic Information** area on the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users. For details, see **Synchronizing IAM Users to MRS**.

NOTE

- When the policy of the user group to which the IAM user belongs changes from MRS ReadOnlyAccess to MRS CommonOperations, MRS FullAccess, or MRS Administrator, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** (System Security Services Daemon) cache of cluster nodes needs time to be updated. Then, submit a job. Otherwise, the job may fail to be submitted.
- When the policy of the user group to which the IAM user belongs changes from MRS CommonOperations, MRS FullAccess, or MRS Administrator to MRS ReadOnlyAccess, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** cache of cluster nodes needs time to be updated.

Step 4 Click the **Jobs** tab.

Step 5 Click **Create**. The **Create Job** page is displayed.

Step 6 Configure job information.

- Set **Type** to **HiveSQL**. Configure other parameters of the HiveSQL job by referring to **Table 4-31**.
- Set **Type** to **Hive Script**. Configure other parameters of the Hive Script job by referring to **Table 4-34**.

Table 4-31 Job configuration information

Parameter	Description
Name	Job name. It contains 1 to 64 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed. **NOTE** You are advised to set different names for different jobs.
SQL Type	Submission type of the SQL statement ● SQL ● Script
SQL Statement	This parameter is valid only when **SQL Type** is set to **SQL**. Enter the SQL statement to be executed, and then click **Check** to check whether the SQL statement is correct. If you want to submit and execute multiple statements at the same time, use semicolons (;) to separate them.

Parameter	Description
SQL File	This parameter is valid only when **SQL Type** is set to **Script**. The path of the SQL file to be executed must meet the following requirements: ● Contains a maximum of 1,023 characters, excluding special characters such as ;|&><'$. The parameter value cannot be empty or full of spaces. ● The path of the program to be executed can be stored in HDFS or OBS. The path varies depending on the file system. – OBS: The path must start with **obs://**. Example: **obs://wordcount/program/*xxx*.jar** – HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. ● For SparkScript and HiveScript, the path must end with **.sql**. For MapReduce, the path must end with **.jar**. For Flink and SparkSubmit, the path must end with **.jar** or **.py**. The **.sql**, **.jar**, and **.py** are case-insensitive. **NOTE** A file path on OBS can start with **obs://**. To submit jobs in this format, you need to configure permissions for accessing OBS. ● If the OBS permission control function is enabled during cluster creation, you can use the **obs://** directory without extra configuration. ● If the OBS permission control function is not enabled or is not supported when you create a cluster, configure the function by following instructions in **Accessing OBS**.
Program Parameter	(Optional) Used to configure optimization parameters such as threads, memory, and vCPUs for the job to optimize resource usage and improve job execution performance. **Table 4-32** describes the common parameters of a running program.
Service Parameter	(Optional) It is used to modify service parameters for the job. The parameter modification applies only to the current job. To make the modification take effect permanently for the cluster, follow instructions in **Configuring Service Parameters**. To add multiple parameters, click ⊕ on the right. To delete a parameter, click **Delete** on the right. **Table 4-33** lists the common service configuration parameters.
Command Reference	Command submitted to the background for execution when a job is submitted.

Table 4-32 Program parameters

Parameter	Description	Example Value
--hiveconf	Hive service configuration, for example, set the execution engine to MapReduce.	Setting the execution engine to MR: **--hiveconf "hive.execution.engine=mr"**
--hivevar	Custom variable, for example, variable ID.	Setting the variable ID: **--hivevar id="123" select * from test where id = ${hivevar:id}**

Table 4-33 Service parameters

Parameter	Description	Example Value
fs.obs.access.key	Key ID for accessing OBS.	-
fs.obs.secret.key	Key corresponding to the key ID for accessing OBS.	-
hive.execution.engine	Engine for running a job.	● mr ● tez

Table 4-34 Job configuration information

Parameter	Description
Name	Job name. It contains 1 to 64 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed. **NOTE** You are advised to set different names for different jobs.

Parameter	Description
Program Path	Path of the program package to be executed. The following requirements must be met: ● Contains a maximum of 1,023 characters, excluding special characters such as ;|&><'$. The parameter value cannot be empty or full of spaces. ● The path of the program to be executed can be stored in HDFS or OBS. The path varies depending on the file system. – OBS: The path must start with **s3a://**. Example: **s3a://wordcount/program/*xxx*.jar** – HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. ● For SparkScript, the path must end with **.sql**. For MapReduce and Spark, the path must end with **.jar**. The **.sql** and **.jar** are case-insensitive.
Parameters	Key parameter for program execution. The parameter is specified by the function of the user's program. MRS is only responsible for loading the parameter. Multiple parameters are separated by space. Configuration method: *Package name*.*Class name* The parameter contains a maximum of 150,000 characters. It cannot contain special characters ;|&><'$, but can be left blank. **NOTE** When entering a parameter containing sensitive information (for example, login password), you can add an at sign (@) before the parameter name to encrypt the parameter value. This prevents the sensitive information from being persisted in plaintext. When you view job information on the MRS management console, the sensitive information is displayed as *****. Example: **username=admin @password=admin_123**
Import From	Path for inputting data Data can be stored in HDFS or OBS. The path varies depending on the file system. ● OBS: The path must start with **s3a://**. ● HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.

Parameter	Description
Export To	Path for outputting data **NOTE** ● When setting this parameter, select **OBS** or **HDFS**. Select a file directory or manually enter a file directory, and click **OK**. ● If you add the **hadoop-mapreduce-examples-x.x.x.jar** sample program or a program similar to **hadoop-mapreduce-examples-x.x.x.jar**, enter a directory that does not exist. Data can be stored in HDFS or OBS. The path varies depending on the file system. ● OBS: The path must start with **s3a://**. ● HDFS: The path must start with **/user**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.
Log Path	Path for storing job logs that record job running status. Data can be stored in HDFS or OBS. The path varies depending on the file system. ● OBS: The path must start with **s3a://**. ● HDFS: The path must start with **/user**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.

Step 7 Confirm job configuration information and click **OK**.

After the job is created, you can manage it.

----End

Submitting a Job in the Background

Step 1 Log in to a Master node. For details, see **Logging In to an ECS**.

Step 2 Run the following command to initialize environment variables:

source /opt/BigData/client/bigdata_env

NOTE

- The default client installation path for MRS 3.x or later is /opt/Bigdata/client, and for versions earlier than MRS 3.x is /opt/client. Configure the path based on site requirements.
- If you use the client to connect to a specific Hive multi-instance in a scenario where multiple Hive instances are installed, run the following command to load the environment variables of the instance. Otherwise, skip this step. For example, load the environment variables of the Hive2 instance.

  source /opt/BigData/client/Hive2/component_env

Step 3 If the Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If the Kerberos authentication is disabled for the current cluster(normal mode), skip this step.

kinit *MRS cluster user* (The user must be in the **hive** user group.)

Step 4 Run the **beeline** command to connect to HiveServer and run tasks.

beeline

For clusters in normal mode, run the following commands. If no component service user is specified, the current OS user is used to log in to the HiveServer.

beeline -n *Component service user*

beeline -f *SQL files* (SQLs in the execution files)

----End

4.5.5 Running a SparkSql Job

You can submit programs developed by yourself to MRS to execute them, and obtain the results. This section describes how to submit a SparkSQL job on the MRS console. SparkSQL jobs are used for data query and analysis. Both SQL statements and scripts are supported. If SQL statements contain sensitive information, use Spark Script to submit them.

Prerequisites

You have uploaded the program packages and data files required for running jobs to OBS or HDFS.

Submitting a Job on the GUI

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters > Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.

Step 3 If Kerberos authentication is enabled for the cluster, perform the following steps. If Kerberos authentication is not enabled for the cluster, skip this step.

In the **Basic Information** area on the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users. For details, see **Synchronizing IAM Users to MRS**.

NOTE

- When the policy of the user group to which the IAM user belongs changes from MRS ReadOnlyAccess to MRS CommonOperations, MRS FullAccess, or MRS Administrator, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** (System Security Services Daemon) cache of cluster nodes needs time to be updated. Then, submit a job. Otherwise, the job may fail to be submitted.
- When the policy of the user group to which the IAM user belongs changes from MRS CommonOperations, MRS FullAccess, or MRS Administrator to MRS ReadOnlyAccess, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** cache of cluster nodes needs time to be updated.

Step 4 Click the **Jobs** tab.

Step 5 For clusters of or later: Click **Create**. On the displayed **Create Job** page, set **Type** to **SparkSql** and configure SparkSql job information by referring to **Table 4-35**.

Table 4-35 Job configuration information

Parameter	Description
Name	Job name. It contains 1 to 64 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed. **NOTE** You are advised to set different names for different jobs.
SQL Type	Submission type of the SQL statement ● SQL ● Script
SQL Statement	This parameter is valid only when **SQL Type** is set to **SQL**. Enter the SQL statement to be executed, and then click **Check** to check whether the SQL statement is correct. If you want to submit and execute multiple statements at the same time, use semicolons (;) to separate them.
SQL File	This parameter is valid only when **SQL Type** is set to **Script**. The path of the SQL file to be executed must meet the following requirements: ● Contains a maximum of 1,023 characters, excluding special characters such as ;|&><'$. The parameter value cannot be empty or full of spaces. ● The path of the program to be executed can be stored in HDFS or OBS. The path varies depending on the file system. – OBS: The path must start with **obs://**. Example: **obs://wordcount/program/*xxx*.jar** – HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. ● For SparkScript and HiveScript, the path must end with **.sql**. For MapReduce, the path must end with **.jar**. For Flink and SparkSubmit, the path must end with **.jar** or **.py**. The **.sql**, **.jar**, and **.py** are case-insensitive. **NOTE** For clusters of or later: A file path on OBS can start with **obs://**. To submit jobs in this format, you need to configure permissions for accessing OBS. ● If the OBS permission control function is enabled during cluster creation, you can use the **obs://** directory without extra configuration. ● If the OBS permission control function is not enabled or is not supported when you create a cluster, configure the function by following instructions in **Accessing OBS**.

Parameter	Description
Program Parameter	(Optional) Used to configure optimization parameters such as threads, memory, and vCPUs for the job to optimize resource usage and improve job execution performance. **Table 4-36** describes the common parameters of a running program.
Service Parameter	(Optional) It is used to modify service parameters for the job. The parameter modification applies only to the current job. To make the modification take effect permanently for the cluster, follow instructions in **Configuring Service Parameters**. To add multiple parameters, click ⊕ on the right. To delete a parameter, click **Delete** on the right. **Table 4-37** lists the common service configuration parameters.
Command Reference	Command submitted to the background for execution when a job is submitted.

Table 4-36 Program parameters

Parameter	Description	Example Value
--conf	Task configuration items to be added.	spark.executor.memory=2G
--driver-memory	Running memory of a driver.	2G
--num-executors	Number of executors to be started.	5
--executor-cores	Number of executor cores.	2
--jars	Additional dependency packages of a task, which is used to add the external dependency packages to the task.	-
--executor-memory	Executor memory.	2G

Table 4-37 Service parameters

Parameter	Description	Example Value
fs.obs.access.key	Key ID for accessing OBS.	-
fs.obs.secret.key	Key corresponding to the key ID for accessing OBS.	-

Step 6 Perform the following operations to create SparkScript and Spark SQL jobs.

- SparkScript job: On the **Jobs** tab page, click **Create**. On the displayed **Create Job** page, set **Type** to **SparkScript** and configure job information by referring to **Table 4-38**.
- Spark SQL job: Click the **Spark SQL** tab, add SQL statements, and submit the SQL statements after a check.

Table 4-38 Job configuration information

Parameter	Description
Name	Job name. It contains 1 to 64 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed. **NOTE** You are advised to set different names for different jobs.
Program Path	Path of the program package to be executed. The following requirements must be met: – Contains a maximum of 1,023 characters, excluding special characters such as ;|&><'$. The parameter value cannot be empty or full of spaces. – The path of the program to be executed can be stored in HDFS or OBS. The path varies depending on the file system. ▪ OBS: The path must start with **s3a://**. Example: **s3a://wordcount/program/*xxx*.jar** ▪ HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. – For SparkScript, the path must end with **.sql**. For MapReduce and Spark, the path must end with **.jar**. The **.sql** and **.jar** are case-insensitive.
Parameters	Key parameter for program execution. The parameter is specified by the function of the user's program. MRS is only responsible for loading the parameter. Multiple parameters are separated by space. Configuration method: *Package name*.*Class name* The parameter contains a maximum of 150,000 characters. It cannot contain special characters ;|&><'$, but can be left blank. **NOTE** When entering a parameter containing sensitive information (for example, login password), you can add an at sign (@) before the parameter name to encrypt the parameter value. This prevents the sensitive information from being persisted in plaintext. When you view job information on the MRS console, the sensitive information is displayed as *. Example: **username=admin @password=admin_123**

Parameter	Description
Import From	Path for inputting data Data can be stored in HDFS. The path varies depending on the file system. – OBS: The path must start with **s3a://**. – HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.
Export To	Path for outputting data **NOTE** – When setting this parameter, select **OBS** or **HDFS**. Select a file directory or manually enter a file directory, and click **OK**. – If you add the **hadoop-mapreduce-examples-x.x.x.jar** sample program or a program similar to **hadoop-mapreduce-examples-x.x.x.jar**, enter a directory that does not exist. Data can be stored in HDFS or OBS. The path varies depending on the file system. – OBS: The path must start with **s3a://**. – HDFS: The path must start with **/user**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.
Log Path	Path for storing job logs that record job running status. Data can be stored in HDFS or OBS. The path varies depending on the file system. – OBS: The path must start with **s3a://**. – HDFS: The path must start with **/user**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.

Step 7 Confirm job configuration information and click **OK**.

After the job is created, you can manage it.

----End

Submitting a Job in the Background

The default client installation path for MRS 3.x or later is /opt/Bigdata/client, and for versions earlier than MRS 3.x is /opt/client. Configure the path based on site requirements.

Step 1 Create a user for submitting jobs. For details, see **Creating a User**.

In this example, a machine-machine user used in the user development scenario has been created, and user groups (**hadoop** and **supergroup**), the primary group (**supergroup**), and role permissions (**System_administrator** and **default**) have been correctly assigned to the user.

Step 2 Download the authentication credential.

- For clusters of MRS 3.*x* or later, log in to FusionInsight Manager and choose **System** > **Permission** > **User**. In the **Operation** column of the newly created user, choose **More** > **Download Authentication Credential**.
- For clusters whose version is earlier than MRS 3.*x*, log in to MRS Manager and choose **System** > **Manage User**. In the **Operation** column of the newly created user, choose **More** > **Download Authentication Credential**.

Step 3 Log in to the node where the Spark client is located, upload the user authentication credential created in 2 to the **/opt** directory of the cluster, and run the following command to decompress the package:

tar –xvf MRSTest _*xxxxxx*_keytab.tar

After the decompression, you obtain the **user.keytab** and **krb5.conf** files.

Step 4 Before performing operations on the cluster, run the following commands:

source /opt/Bigdata/client/bigdata_env

cd $SPARK_HOME

Step 5 Open the **spark-sql** CLI and run the following SQL statement:

./bin/spark-sql --conf spark.yarn.principal=MRSTest --conf spark.yarn.keytab=/opt/user.keytab

To execute the SQL file, you need to upload the SQL file (for example, to the **/opt/** directory). After the file is uploaded, run the following command:

./bin/spark-sql --conf spark.yarn.principal=MRSTest --conf spark.yarn.keytab=/opt/user.keytab -f /opt/script.sql

----End

4.5.6 Running a Flink Job

You can submit programs developed by yourself to MRS to execute them, and obtain the results. This section describes how to submit a Flink job on the MRS management console. Flink jobs are used to submit JAR programs to process streaming data.

Prerequisites

You have uploaded the program packages and data files required for running jobs to OBS or HDFS.

Submitting a Job on the GUI

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters > Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.

Step 3 If Kerberos authentication is enabled for the cluster, perform the following steps. If Kerberos authentication is not enabled for the cluster, skip this step.

In the **Basic Information** area on the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users. For details, see **Synchronizing IAM Users to MRS**.

NOTE

- When the policy of the user group to which the IAM user belongs changes from MRS ReadOnlyAccess to MRS CommonOperations, MRS FullAccess, or MRS Administrator, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** (System Security Services Daemon) cache of cluster nodes needs time to be updated. Then, submit a job. Otherwise, the job may fail to be submitted.
- When the policy of the user group to which the IAM user belongs changes from MRS CommonOperations, MRS FullAccess, or MRS Administrator to MRS ReadOnlyAccess, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** cache of cluster nodes needs time to be updated.

Step 4 Click the **Jobs** tab.

Step 5 Click **Create**. The **Create Job** page is displayed.

Step 6 Set **Type** to **Flink**. Configure Flink job information by referring to **Table 4-39**.

Table 4-39 Job configuration information

Parameter	Description
Name	Job name. It contains 1 to 64 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed. **NOTE** You are advised to set different names for different jobs.
Program Path	Path of the program package to be executed. The following requirements must be met: ● Contains a maximum of 1,023 characters, excluding special characters such as ;|&><'$. The parameter value cannot be empty or full of spaces. ● The path of the program to be executed can be stored in HDFS or OBS. The path varies depending on the file system. – OBS: The path must start with **obs://**. Example: **obs://wordcount/program/*xxx*.jar** – HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**.
Program Parameter	(Optional) Used to configure optimization parameters such as threads, memory, and vCPUs for the job to optimize resource usage and improve job execution performance. **Table 4-40** describes the common parameters of a running program.

Parameter	Description
Parameters	(Optional) Key parameter for program execution. The parameter is specified by the function of the user's program. MRS is only responsible for loading the parameter. Multiple parameters are separated by space. The parameter contains a maximum of 150,000 characters. It cannot contain special characters ;|&><'$, but can be left blank. **CAUTION** If you enter a parameter with sensitive information (such as the login password), the parameter may be exposed in the job details display and log printing. Exercise caution when performing this operation.
Service Parameter	(Optional) It is used to modify service parameters for the job. The parameter modification applies only to the current job. To make the modification take effect permanently for the cluster, follow instructions in **Configuring Service Parameters**. To add multiple parameters, click ⊕ on the right. To delete a parameter, click **Delete** on the right. **Table 4-41** describes the common parameters of a service.
Command Reference	Command submitted to the background for execution when a job is submitted.

Table 4-40 Program parameters

Parameter	Description	Example Value
-ytm	Memory size of each TaskManager container. (Optional unit. The unit is MB by default.)	1024
-yjm	Memory size of JobManager container. (Optional unit. The unit is MB by default.)	1024
-yn	Number of Yarn containers allocated to applications. The value is the same as the number of TaskManagers.	2
-ys	Number of TaskManager cores.	2
-ynm	Custom name of an application on Yarn.	test
-c	Class of the program entry point (for example, the **main** or **getPlan()** method). This parameter is required only when the JAR file does not specify the class of its manifest.	com.bigdata.mrs.test

NOTE

For MRS 3.x or later, the **-yn** parameter is not supported.

Table 4-41 Service parameters

Parameter	Description	Example Value
fs.obs.access.key	Key ID for accessing OBS.	-
fs.obs.secret.key	Key corresponding to the key ID for accessing OBS.	-

Step 7 Confirm job configuration information and click **OK**.

After the job is created, you can manage it.

----End

Submitting a Job in the Background

The default client installation path for MRS 3.x or later is /opt/Bigdata/client, and for versions earlier than MRS 3.x is /opt/client. Configure the path based on site requirements.

Step 1 Log in to the MRS client.

Step 2 Run the following command to initialize environment variables:

source /opt/Bigdata/client/bigdata_env

Step 3 If Kerberos authentication is enabled for the cluster, perform the following steps. If Kerberos authentication is not enabled for the cluster, skip this step.

1. Prepare a user for submitting Flink jobs.
2. Log in to Manager as the newly created user.
   - For MRS 3.x earlier: Log in to Manager of the cluster. Choose **System** > **Manage User**. In the **Operation** column of the row that contains the added user, choose **More** > **Download authentication credential** to locate the row that contains the user.
   - For MRS 3.*x* or later: Log in to Manager of the cluster. Choose **System** > **Permission** > **Manage User**. On the displayed page, locate the row that contains the added user, click **More** in the **Operation** column, and select **Download authentication credential**.
3. Decompress the downloaded authentication credential package and copy the **user.keytab** file to the client node, for example, to the **/opt/Bigdata/client/Flink/flink/conf** directory on the client node. If the client is installed on a node outside the cluster, copy the **krb5.conf** file to the **/etc/** directory on this node.
4. For MRS 3.x or later: In security mode, add the service IP address of the node where the client is installed and floating IP address of Manager to the **jobmanager.web.allow-access-address** configuration item in the **/opt/Bigdata/client/Flink/flink/conf/flink-conf.yaml** file.

5. Run the following commands to configure security authentication by adding the **keytab** path and username to the **/opt/Bigdata/client/Flink/flink/conf/flink-conf.yaml** configuration file.

   security.kerberos.login.keytab: *<user.keytab file path>*

   security.kerberos.login.principal: *<Username>*

   Example:

   security.kerberos.login.keytab: /opt/Bigdata/client/Flink/flink/conf/user.keytab

   security.kerberos.login.principal: test

6. Run the following command to perform security hardening in the **bin** directory of the Flink client. Set password to a new password for submitting jobs.

   sh generate_keystore.sh <*password*>

   This script automatically replaces the SSL value in the **/opt/Bigdata/client/Flink/flink/conf/flink-conf.yaml** file. For MRS 3.x or earlier, external SSL is disabled by default in security clusters. To enable external SSL, run this script again after configuration. The configuration parameters do not exist in the default Flink configuration of MRS, if you enable SSL for external connections, you need to add the parameters listed in **Table 4-42**.

Table 4-42 Parameter description

Parameter	Example Value	Description
security.ssl.rest.enabled	true	Switch to enable external SSL.
security.ssl.rest.keystore	${path}/flink.keystore	Path for storing **keystore**.
security.ssl.rest.keystore-password	123456	Password of the **keystore**. **123456** indicates a user-defined password is required.
security.ssl.rest.key-password	123456	Password of the SSL key. **123456** indicates a user-defined password is required.
security.ssl.rest.truststore	${path}/flink.truststore	Path for storing the **truststore**.
security.ssl.rest.truststore-password	123456	Password of the **truststore**. **123456** indicates a user-defined password is required.

NOTE

- For MRS 3.x or earlier: The **generate_keystore.sh** script is automatically generated.
- Perform **authentication and encryption**. The generated **flink.keystore**, **flink.truststore**, and **security.cookie** files are automatically filled in the corresponding configuration items in **flink-conf.yaml**.
- For MRS 3.*x* or later: You can obtain the values of **security.ssl.key-password**, **security.ssl.keystore-password**, and **security.ssl.truststore-password** using the Manager plaintext encryption API by running the following command:

  curl -k -i -u <*user name*>:<*password*> -X POST -HContent-type:application/json -d '{"plainText":"<*password*>"}' 'https://*x.x.x.x*:28443/web/api/v2/tools/encrypt'; In the preceding command, <*password*> must be the same as the password used for issuing the certificate, and *x.x.x.x* indicates the floating IP address of Manager in the cluster.

7. Configure paths for the client to access the **flink.keystore** and **flink.truststore** files.
   - Absolute path: After the script is executed, the file path of **flink.keystore** and **flink.truststore** is automatically set to the absolute path **opt/Bigdata/client/Flink/flink/conf/** in the **flink-conf.yaml** file. In this case, you need to move the **flink.keystore** and **flink.truststore** files from the **conf** directory to this absolute path on the Flink client and Yarn nodes.
   - Relative path: Perform the following steps to set the file path of **flink.keystore** and **flink.truststore** to the relative path and ensure that the directory where the Flink client command is executed can directly access the relative paths.
     i. In the **/opt/Bigdata/client/Flink/flink/conf/**directory, create a new directory, for example, **ssl**.
     ii. Move the **flink.keystore** and **flink.truststore** file to the /**opt/Bigdata/client/Flink/flink/conf/ssl/** directory.
     iii. For MRS 3.*x* or later: Change the values of the following parameters in the **flink-conf.yaml** file to relative paths:
     ```
     security.ssl.keystore: ssl/flink.keystore
     security.ssl.truststore: ssl/flink.truststore
     ```
     iv. For MRS 3.x or earlier: Change the values of the following parameters in the **flink-conf.yaml** file to relative paths:
     ```
     security.ssl.internal.keystore: ssl/flink.keystore
     security.ssl.internal.truststore: ssl/flink.truststore
     ```
8. If the client is installed on a node outside the cluster, add the following configuration to the configuration file (for example, **/opt/Bigdata/client/Flink/fink/conf/flink-conf.yaml**). Replace ***xx.xx.xxx.xxx*** with the IP address of the node where the client resides.
   ```
   web.access-control-allow-origin: xx.xx.xxx.xxx
   jobmanager.web.allow-access-address: xx.xx.xxx.xxx
   ```

Step 4 Run a wordcount job.

- Normal cluster (Kerberos authentication disabled)
  - Run the following commands to start a session and submit a job in the session:
  ```
  yarn-session.sh -nm "session-name"
  flink run /opt/Bigdata/client/Flink/flink/examples/streaming/WordCount.jar
  ```
  - Run the following command to submit a single job on Yarn:
  ```
  flink run -m yarn-cluster /opt/Bigdata/client/Flink/flink/examples/streaming/
  WordCount.jar
  ```

- Security cluster (Kerberos authentication enabled)
  - If the **flink.keystore** and **flink.truststore** file are stored in the absolute path:
    - Run the following commands to start a session and submit a job in the session:
      ```
      yarn-session.sh -nm "session-name"
      flink run /opt/Bigdata/client/Flink/flink/examples/streaming/WordCount.jar
      ```
    - Run the following command to submit a single job on Yarn:
      ```
      flink run -m yarn-cluster /opt/Bigdata/client/Flink/flink/examples/streaming/WordCount.jar
      ```
  - If the **flink.keystore** and **flink.truststore** file are stored in the relative path:
    - In the same directory of SSL, run the following command to start a session and submit jobs in the session. The SSL directory is a relative path. For example, if the SSL directory is **opt/Bigdata/client/Flink/flink/conf/**, then run the following command in this directory:
      ```
      yarn-session.sh -t ssl/ -nm "session-name"
      flink run /opt/Bigdata/client/Flink/flink/examples/streaming/WordCount.jar
      ```
    - Run the following command to submit a single job on Yarn:
      ```
      flink run -m yarn-cluster -yt ssl/ /opt/Bigdata/client/Flink/flink/examples/streaming/WordCount.jar
      ```

----End

4.5.7 Running a Kafka Job

You can submit programs developed by yourself to MRS to execute them, and obtain the results. This topic describes how to generate and consume messages in a Kafka topic.

Currently, Kafka jobs cannot be submitted on the GUI. You can submit them in the background.

Submitting a Job in the Background

Query the instance addresses of ZooKeeper and Kafka, and then run the Kafka job.

Querying the Instance Address (3.x)

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters** > **Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.

Step 3 Go to the FusionInsight Manager page. For details, see **Accessing FusionInsight Manager (MRS 3.x or Later)**. On MRS Manager, choose **Services** > **ZooKeeper** > **Instance** to query the IP addresses of ZooKeeper instances. Record any IP address of a ZooKeeper instance.

Step 4 Choose **Services** > **Kafka** > **Instance** to query the IP addresses of Kafka instances. Record any IP address of a Kafka instance.

----End

Querying the Instance Address (Versions Earlier Than 3.x)

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters** > **Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.

Step 3 On the MRS cluster details page, choose **Components > ZooKeeper > Instance** to query the IP addresses of ZooKeeper instances. Record any IP address of a ZooKeeper instance.

Step 4 Choose **Components > Kafka > Instance** to query the IP addresses of Kafka instances. Record any IP address of a Kafka instance.

----End

Running a Kafka Job

The default client installation path for MRS 3.x or later is /opt/Bigdata/client, and for versions earlier than MRS 3.x is /opt/client. Configure the path based on site requirements.

Step 1 Log in to the Master2 node. For details, see **Logging In to an ECS**.

Step 2 Run the following command to initialize environment variables:

source /opt/Bigdata/client/bigdata_env

Step 3 If the Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If the Kerberos authentication is disabled for the current cluster, skip this step.

kinit *MRS cluster user*

Example: **kinit admin**

Step 4 Run the following command to create a Kafka topic:

kafka-topics.sh --create --zookeeper <IP address of the ZooKeeper role instance:2181/kafka> --partitions 2 --replication-factor 2 --topic <Topic name>

Step 5 Produce messages in a topic test.

Run the following command: **kafka-console-producer.sh --broker-list <IP address of the Kafka role instance:9092> --topic <Topic name> --producer.config /opt/Bigdata/client/Kafka/kafka/config/producer.properties**.

Input specified information as the messages produced by the producer and then press **Enter** to send the messages. To end message production, press **Ctrl+C** to exit.

Step 6 Consume messages in the topic test.

kafka-console-consumer.sh --topic <Topic name> --bootstrap-server <Kafka role instance IP:210079092> --consumer.config /opt/Bigdata/client/Kafka/kafka/config/consumer.properties

NOTE

If Kerberos authentication is enabled in the cluster, change the port number 9092 to 21007 when running the preceding two commands. For details, see **List of Open Source Component Ports**.

----End

4.5.8 Viewing Job Configuration and Logs

This section describes how to view job configuration and logs.

Background

- You can view configuration information of all jobs.
- You can only view logs of running jobs.

  Because logs of Spark SQL and DistCp jobs are not in the background, you cannot view logs of running Spark SQL and DistCp jobs.

Procedure

Step 1 Log in to the MRS management console.

Step 2 Choose **Clusters** > **Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.

Step 3 Click **Jobs**.

Step 4 In the **Operation** column of the job to be viewed, click **View Details**.

In the **View Details** window that is displayed, configuration of the selected job is displayed.

Step 5 Select a running job, and click **View Log** in the **Operation** column.

In the new page that is displayed, real-time log information of the job is displayed.

Each tenant can submit and view 10 jobs concurrently.

----End

4.5.9 Stopping a Job

This section describes how to stop running MRS jobs.

Background

You cannot stop Spark SQL jobs. After a job is stopped, its status changes to **Terminated** and the job cannot be executed again.

Procedure

Step 1 Log in to the MRS management console.

Step 2 Choose **Clusters** > **Active Clusters**, select a running cluster, and click its name.

The cluster details page is displayed.

Step 3 Click **Jobs**.

Step 4 Select a running job, and choose **More > Stop** in the **Operation** column.

The job status changes from **Running** to **Terminated**.

----End

4.5.10 Deleting a Job

This section describes how to delete an MRS job. After a job is executed, you can delete it if you do not need to view its information.

Background

Jobs can be deleted one after another or in a batch. A deleted job cannot be restored. Therefore, exercise caution when deleting a job.

Procedure

Step 1 Log in to the MRS management console.

Step 2 Choose **Clusters** > **Active Clusters**, select a running cluster, and click its name.

The cluster details page is displayed.

Step 3 Click **Jobs**.

Step 4 Choose **More > Delete** from the **Operation** in the row of the target job to be deleted.

In this step, you can only delete one job only.

Step 5 If you select multiple jobs and click **Delete** on the upper left of the job list.

You can delete one, multiple, or all jobs.

----End

4.5.11 Configuring Job Notification Rules

MRS uses SMN to offer a publish/subscribe model to achieve one-to-multiple message subscriptions and notifications in a variety of message types (SMSs and emails). You can configure job notification rules to receive notifications immediately upon a job execution success or failure.

Procedure

Step 1 Log in to the management console.

Step 2 Click **Service List**. Under **Management & Governance**, click **Simple Message Notification**.

Step 3 Create a topic and add subscriptions to the topic. For details, see **Configuring Message Notification**.

Step 4 Go to the MRS management console, and click the cluster name to go to the cluster details page.

Step 5 Click the **Alarms** tab, and choose **Notification Rules** > **Add Notification Rule**.

Step 6 Configure a notification rule for sending job execution results to subscribers.

Table 4-43 Parameters of adding a notification rule

Parameter	Description
Rule Name	User-defined notification rule name. Only digits, letters, hyphens (-), and underscores (_) are allowed.
Message Notification	If you enable this function, subscription messages will be sent to subscribers.
Topic Name	Select an existing topic or click **Create Topic** to create a topic.
Notification Type	Select **Event**.
Subscription Items	1. Click ⊞ next to **Suggestion**. 2. Click ⊞ next to **Manager**. 3. Select **Job Running Succeeded** and **Job Running Failed**.

----End

4.6 Component Management

4.6.1 Object Management

MRS contains different types of basic objects. **Table 4-44** describes these objects.

Table 4-44 MRS basic object overview

Object	Description	Example
Service	Function set that can complete specific business.	KrbServer service and LdapServer service
Service instance	Specific instance of a service, usually called service.	KrbServer service
Service role	Function entity that forms a complete service, usually called role.	KrbServer is composed of the KerberosAdmin role and KerberosServer role.
Role instance	Specific instance of a service role running on a host.	KerberosAdmin that is running on Host2 and KerberosServer that is running on Host3

Object	Description	Example
Host	An ECS running Linux OS.	Host1 to Host5
Rack	Physical entity that contains multiple hosts connecting to the same switch.	Rack1 contains Host1 to Host5.
Cluster	Logical entity that consists of multiple hosts and provides various services.	Cluster1 cluster consists of five hosts (Host1 to Host5) and provides services such as KrbServer and LdapServer.

4.6.2 Viewing Configuration

On MRS, you can view the configuration of services (including roles) and role instances.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

- Query service configuration.
  a. On the MRS cluster details page, click **Components**.
  b. Select the target service from the service list.
  c. Click **Service Configuration**.
  d. Switch **Basic** to **All**. All configuration parameters of the service are displayed in the navigation tree. The service name and role names are displayed from upper to lower in the navigation tree.
  e. In the navigation tree, select a specified parameter and change its value. You can also enter the parameter name in the **Search** box to search for the parameter and view the result.

     The parameters under the service nodes and role nodes are service configuration parameters and role configuration parameters respectively.
  f. Select **Non-default** from the **--Select--** drop-down list. The parameters whose values are not default values are displayed.
- Query role instance configurations.
  a. On the MRS cluster details page, click **Components**.
  b. Select the target service from the service list.
  c. Click the **Instances** tab.
  d. Click the target role instance from the role instance list.
  e. Click **Instance Configuration**.
  f. Switch **Basic** to **All** on the right of the page. All configuration parameters of the role instance are displayed in the navigation tree.

g. In the navigation tree, select a specified parameter and change its value. You can also enter the parameter name in the **Search** box to search for the parameter and view the result.

h. Select **Non-default** from the **--Select--** drop-down list. The parameters whose values are not default values are displayed.

4.6.3 Managing Services

You can perform the following operations on MRS:

- Start the service in the **Stopped**, **Stop Failed**, or **Failed to Start** state to use the service.
- Stop the services or stop abnormal services.
- Restart abnormal services or configure expired services to restore or enable the services.

Prerequisites

- You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Impact on the System

- The stateful component cannot be added to the task node group.

Starting, Stopping, and Restarting a Service

Step 1 On the MRS cluster details page, click **Components**.

Step 2 Locate the row that contains the target service, **Start**, **Stop**, and **Restart** to start, stop, or restart the service.

Services are interrelated. If a service is started, stopped, and restarted, services dependent on it will be affected.

The services will be affected in the following ways:

- If a service is to be started, the lower-layer services dependent on it must be started first.
- If a service is stopped, the upper-layer services dependent on it are unavailable.
- If a service is restarted, the running upper-layer services dependent on it must be restarted.

----End

4.6.4 Configuring Service Parameters

On the MRS console, you can view and modify the default service configurations based on site requirements and export or import the configurations.

Impact on the System

- You need to download and update the client configuration files after configuring HBase, HDFS, Hive, Spark, Yarn, and MapReduce service properties.
- The parameters of DBService cannot be modified when only one DBService role instance exists in the cluster.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Modifying Service Parameters

1. On the MRS cluster details page, click **Components**.
2. Select the target service from the service list.
3. Click **Service Configuration**.
4. Switch **Basic** to **All**. All configuration parameters of the service are displayed in the navigation tree. The service name and role names are displayed from upper to lower in the navigation tree.
5. In the navigation tree, select a specified parameter and change its value. You can also enter the parameter name in the **Search** box to search for the parameter and view the result.

   If you want to cancel the modification of a parameter value, click ↻ to restore it.
6. Click **Save Configuration**, select **Restart the affected services or instances**, and click **OK**.

   NOTE

   To update the queue configuration of Yarn without restarting service, choose **More** > **Refresh Queue** on the **Service Status** tab page to update the queue for the configuration to take effect.

4.6.5 Configuring Customized Service Parameters

Each component of MRS supports all open-source parameters. MRS supports the modification of some parameters for key application scenarios. Some component clients may not include all parameters with open-source features. To modify the component parameters that are not directly supported by MRS, you can add new parameters for components by using the configuration customization function on MRS. Newly added parameters are saved in component configuration files and take effect after restart.

Impact on the System

- After the service attributes are configured, the service needs to be restarted. The service cannot be accessed during restart.
- You need to download and update the client configuration files after configuring HBase, HDFS, Hive, Spark, Yarn, and MapReduce service properties.

Prerequisites

- You have understood the meanings of parameters to be added, configuration files that have taken effect, and the impact on components.
- You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS cluster details page, click **Components**.

Step 2 Select the target service from the service list.

Step 3 Click **Service Configuration**.

Step 4 In the configuration type drop-down box on the right side, switch **Basic** to **All**.

Step 5 In the navigation tree, select **Customization**. The customized parameters of the current component are displayed on MRS.

The configuration files that save the newly added customized parameters are displayed in the **Parameter File** column. Different configuration files may have same open-source parameters. After the parameters in different files are set to different values, whether the configuration takes effect depends on the loading sequence of the configuration files by components. You can customize parameters for services and roles as required. Adding customized parameters for a single role instance is not supported.

Step 6 Based on the configuration files and parameter functions, locate the row where a specified parameter resides, enter the parameter name supported by the component in the **Parameter** column and enter the parameter value in the **Value** column.

- You can click ⊕ or ⊗ to add or delete a custom parameter. You can delete a customized parameter only after you click ⊕ for the first time.
- If you want to cancel the modification of a parameter value, click ↻ to restore it.

Step 7 Click **Save Configuration**, select **Restart the affected services or instances**, and click **OK**.

----End

Task Example

Configuring Customized Hive Parameters

Hive depends on HDFS. By default, Hive accesses the HDFS client. The configuration parameters to take effect are controlled by HDFS in a unified manner. For example, the HDFS parameter **ipc.client.rpc.timeout** affects the RPC timeout period for all clients to connect to the HDFS server. If you need to modify the timeout period for Hive to connect to HDFS, you can use the configuration customization function. After this parameter is added to the **core-site.xml** file of Hive, this parameter can be identified by the Hive service and its configuration overwrites the parameter configuration in HDFS.

Step 1 On the MRS cluster details page, click **Components**.

Step 2 Choose **Hive** > **Service Configuration**.

Step 3 In the configuration type drop-down box on the right side, switch **Basic** to **All**.

Step 4 In the navigation tree on the left, select **Customization** for the Hive service. The system displays the customized service parameters supported by Hive.

Step 5 In **core-site.xml**, locate the row that contains the **core.site.customized.configs** parameter, enter **ipc.client.rpc.timeout** in the **Parameter** column, and enter a new value in the **Value** column, for example, **150000**. The unit is millisecond.

Step 6 Click **Save Configuration**, select **Restart the affected services or instances**, and click **OK**.

Operation successful is displayed. Click **Finish**. The service is started successfully.

----End

4.6.6 Synchronizing Service Configuration

Scenario

If **Configuration Status** of some services is **Configuration expired** or **Configuration failed**, synchronize configuration for the cluster or service to restore its configuration status. If all services in the cluster are in the **Configuration failed** state, synchronize the cluster configuration with the background configuration.

Impact on the System

After synchronizing service configurations, you need to restart the services whose configurations have expired. These services are unavailable during restart.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS cluster details page, click **Components**.

Step 2 Select the target service from the service list.

Step 3 On the Service Status tab page, choose **More** > **Synchronize Configuration**.

Step 4 In the dialog box that is displayed, select **Restart the service or instances whose configurations have expired** and click **Yes** to restart the service.

----End

4.6.7 Managing Role Instances

Scenario

You can start a role instance that is in the **Stopped**, **Failed to stop** or **Failed to start** status, stop an unused or abnormal role instance or restart an abnormal role instance to recover its functions.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS cluster details page, click **Components**.

Step 2 Select the target service from the service list.

Step 3 Click the **Instances** tab.

Step 4 Select the check box on the left of the target role instance.

Step 5 Click **More**, select operations such as **Start Instance**, **Stop Instance**, **Restart Instance**, **Rolling-restart Instance**, or **Delete Instance** based on site requirements.

----End

4.6.8 Configuring Role Instance Parameters

Scenario

You can view and modify default role instance configuration on MRS based on site requirements. The configurations can be imported and exported.

Impact on the System

You need to download and update the client configuration files after configuring HBase, HDFS, Hive, Spark, Yarn, and MapReduce service properties.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Modifying Role Instance Parameters

1. On the MRS cluster details page, click **Components**.
2. Select the target service from the service list.
3. Click the **Instances** tab.
4. Click the target role instance from the role instance list.

5. Click the **Instance Configuration** tab.
6. Switch **Basic** to **All** from the drop-down list on the right of the page. All configuration parameters of the role instance are displayed in the navigation tree.
7. In the navigation tree, select a specified parameter and change its value. You can also enter the parameter name in the **Search** box to search for the parameter and view the result.

   If you want to cancel the modification of a parameter value, click ↻ to restore it.
8. Click **Save Configuration**, select **Restart the affected services or instances**, and click **OK**.

4.6.9 Synchronizing Role Instance Configuration

Scenario

When **Configuration Status** of a role instance is **Configuration expired** or **Configuration failed**, you can synchronize the configuration data of the role instance with the background configuration.

Impact on the System

After synchronizing a role instance configuration, you need to restart the role instance whose configuration has expired. The role instance is unavailable during restart.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS cluster details page, click **Components**.

Step 2 Select a service name.

Step 3 Click the **Instances** tab.

Step 4 Click the target role instance from the role instance list.

Step 5 Choose **More** > **Synchronize Configuration** above the role instance status and indicator information.

Step 6 In the dialog box that is displayed, select **Restart the service or instances whose configurations have expired** and click **Yes** to restart the role instance.

----End

4.6.10 Decommissioning and Recommissioning a Role Instance

Scenario

If a Core or Task node is faulty, the cluster status may be displayed as **Abnormal**. In an MRS cluster, data can be stored on different Core nodes. You can decommission the specified role instance on MRS to stop the role instance from providing services. After fault rectification, you can recommission the role instance.

The following role instances can be decommissioned or recommissioned:

- DataNode role instance on HDFS
- NodeManager role instance on Yarn
- RegionServer role instance on HBase
- ClickHouseServer role instance on ClickHouse
- Broker role instance on Kafka

Restrictions:

- If the number of the DataNodes is less than or equal to that of HDFS copies, decommissioning cannot be performed. If the number of HDFS copies is three and the number of DataNodes is less than four in the system, decommissioning cannot be performed. In this case, an error will be reported and force MRS to exit the decommissioning 30 minutes after MRS attempts to perform the decommissioning.
- If the number of Kafka Broker instances is less than or equal to that of Kafka copies, decommissioning cannot be performed. For example, if the number of Kafka copies is two and the number of nodes is less than three in the system, decommissioning cannot be performed. Instance decommissioning will fail and exit.
- If a role instance is out of service, you must recommission the instance to start it before using it again.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS cluster details page, click **Components**.

Step 2 Click a service in the service list.

Step 3 Click the **Instances** tab.

Step 4 Select an instance.

Step 5 Choose **More** > **Decommission** or **Recommission** to perform the corresponding operation.

NOTE

During the instance decommissioning, if the service corresponding to the instance is restarted in the cluster using another browser, MRS displays a message indicating that the instance decommissioning is stopped, but the **Operating Status** of the instance is displayed as **Started**. In this case, the instance has been decommissioned on the background. You need to decommission the instance again to synchronize the operating status.

----End

4.6.11 Starting and Stopping a Cluster

A cluster is a collection of service components. You can start or stop all services in a cluster.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

On the cluster details page, choose **Management Operations** > **Start All Components** or **Stop All Components** in the upper right corner to perform the required operation.

4.6.12 Synchronizing Cluster Configuration

Scenario

If **Configuration Status** of all services or some services is **Configuration expired** or **Configuration failed**, synchronize configuration for the cluster or service to restore its configuration status.

- If all services in the cluster are in the **Configuration failed** status, synchronize the cluster configuration with the background configuration.
- If all services in the cluster are in the **Configuration failed** status, synchronize the service configuration with the background configuration.

  NOTE

  In **MRS 3.x**, you cannot perform operations in this section on the management console.

Impact on the System

After synchronizing cluster configurations, you need to restart the services whose configurations have expired. These services are unavailable during restart.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the cluster details page, choose **Configuration** > **Synchronize Configuration** in the upper right corner.

Step 2 In the displayed dialog box, select **Restart services and instances whose configuration have expired**, and click **OK** to restart the service whose configuration has expired.

When **Operation successful** is displayed, click **Finish**. The cluster is started successfully.

----End

4.6.13 Exporting Cluster Configuration

Scenario

You can export all configuration data of a cluster using MRS to meet site requirements. The exported configuration data is used to rapidly update service configuration.

NOTE

In **MRS 3.x**, you cannot perform operations in this section on the management console.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

On the cluster details page, choose **Configuration** > **Export Cluster Configuration** in the upper right corner.

The exported file is used to update service configurations. For details, see **Importing Service Configuration Parameters** in **Configuring Service Parameters**.

4.6.14 Performing Rolling Restart

After modifying the configuration items of a big data component, you need to restart the corresponding service to make new configurations take effect. If you use a normal restart mode, all services or instances are restarted concurrently, which may cause service interruption. To ensure that services are not affected during service restart, you can restart services or instances in batches by rolling restart. For instances in active/standby mode, a standby instance is restarted first and then an active instance is restarted. Rolling restart takes longer than normal restart.

Table 4-45 provides services and instances that support or do not support rolling restart in the MRS cluster.

Table 4-45 Services and instances that support or do not support rolling restart

Service	Instance	Whether to Support Rolling Restart
HDFS	NameNode	Yes
	Zkfc	
	JournalNode	
	HttpFS	
	DataNode	
Yarn	ResourceManager	Yes
	NodeManager	
Hive	MetaStore	Yes
	WebHCat	
	HiveServer	
Mapreduce	JobHistoryServer	Yes
HBase	HMaster	Yes
	RegionServer	
	ThriftServer	
	RESTServer	
Spark	JobHistory	Yes
	JDBCServer	
	SparkResource	No
Hue	Hue	No
Tez	TezUI	No
Loader	Sqoop	No
Zookeeper	Quorumpeer	Yes
Kafka	Broker	Yes
	MirrorMaker	No
Flume	Flume	Yes
	MonitorServer	
Storm	Nimbus	Yes
	UI	
	Supervisor	

Service	Instance	Whether to Support Rolling Restart
	Logviewer	

Restrictions

- Perform a rolling restart during off-peak hours.
  - Otherwise, a rolling restart failure may occur. For example, if the throughput of Kafka is high (over 100 MB/s) during the Kafka rolling restart, the Kafka rolling restart may fail.
  - For example, if the requests per second of each RegionServer on the native interface exceed 10,000 during the HBase rolling restart, you need to increase the number of handles to prevent a RegionServer restart failure caused by heavy loads during the restart.
- Before the restart, check the number of current requests of HBase. If the number of requests of each RegionServer on the native interface exceeds 10,000, increase the number of handles to prevent a failure.
- If the number of Core nodes in a cluster is less than six, services may be affected for a short period of time.
- Preferentially perform a rolling instance or service restart and select **Only restart instances whose configurations have expired**.

Performing a Rolling Service Restart

Step 1 Choose **Clusters** > **Active Clusters** and click a cluster name to go to the cluster details page.

Step 2 Click **Components** and select a service for which you want to perform a rolling restart.

Step 3 On the **Service Status** tab page, click **More** and select **Rolling-restart Service**.

Step 4 The **Rolling-restart Service** page is displayed. Select **Only restart instances whose configurations have expired** and click **OK** to perform rolling restart for the service.

Step 5 After the rolling restart task is complete, click **Finish**.

----End

Performing a Rolling Instance Restart

Step 1 Choose **Clusters** > **Active Clusters** and click a cluster name to go to the cluster details page.

Step 2 Click **Components** and select a service for which you want to perform a rolling restart.

Step 3 On the **Instance** tab page, select the instance to be restarted. Click **More** and select **Rolling-restart Instance**.

Step 4 After you enter the administrator password, the **Rolling-restart Instance** page is displayed. Select **Only restart instances whose configurations have expired** and click **OK** to perform rolling restart for the instance.

Step 5 After the rolling restart task is complete, click **Finish**.

----End

Perform a Rolling Cluster Restart

Step 1 Choose **Clusters** > **Active Clusters** and click a cluster name to go to the cluster details page.

Step 2 In the upper right corner of the page, choose **Management Operations** > **Perform Rolling Cluster Restart**.

Step 3 The **Rolling-restart Cluster** page is displayed. Select **Only restart instances whose configurations have expired** and click **OK** to perform rolling restart for the cluster.

Step 4 After the rolling restart task is complete, click **Finish**.

----End

Rolling Restart Parameter Description

Table 4-46 describes rolling restart parameters.

Table 4-46 Rolling restart parameter description

Parameter	Description
Only restart instances whose configurations have expired	Specifies whether to restart only the modified instances in a cluster.
Data Node Instances to Be Batch Restarted	Specifies the number of instances that are restarted in each batch when the batch rolling restart strategy is used. The default value is **1**. The value ranges from 1 to 20. This parameter is valid only for data nodes.
Batch Interval	Specifies the interval between two batches of instances for rolling restart. The default value is **0**. The value ranges from 0 to 2147483647. The unit is second. Note: Setting the batch interval parameter can increase the stability of the big data component process during the rolling restart. You are advised to set this parameter to a non-default value, for example, 10.
Batch Fault Tolerance Threshold	Specifies the tolerance times when the rolling restart of instances fails to be executed in batches. The default value is **0**, which indicates that the rolling restart task ends after any batch of instances fails to be restarted. The value ranges from 0 to 2147483647.

Procedure in a Typical Scenario

Step 1 Choose **Clusters** > **Active Clusters** and click a cluster name to go to the cluster details page.

Step 2 Click **Components** and select **HBase**. The **HBase** service page is displayed.

Step 3 Click the **Service Configuration** tab, and modify an HBase parameter. After the following dialog box is displayed, click **OK** to save the configurations.

NOTE

Do not select **Restart the affected services or instances**. This option indicates a normal restart. If you select this option, all services or instances will be restarted, which may cause service interruption.

Step 4 After saving the configurations, click **Finish**.

Step 5 Click the **Service Status** tab.

Step 6 On the **Service Status** tab page, click **More** and select **Rolling-restart Service**.

Step 7 After you enter the administrator password, the **Rolling-restart Service** page is displayed. Select **Only restart instances whose configurations have expired** and click **OK** to perform rolling restart.

Step 8 After the rolling restart task is complete, click **Finish**.

----End

4.7 Alarm Management

4.7.1 Viewing the Alarm List

The alarm list displays all alarms in the MRS cluster. The MRS page displays the alarms that need to be handled in a timely manner and the events.

On the MRS management console, you can only query basic information about uncleared MRS alarms on the **Alarms** tab page. For details about how to view alarm details or manage alarms, see **Viewing and Manually Clearing an Alarm**.

Alarms are listed in chronological order by default in the alarm list, with the most recent alarms displayed at the top.

Table 4-47 describes various fields in an alarm.

Table 4-47 Alarm description

Parameter	Description
Alarm ID	ID of an alarm.
Alarm Name	Name of an alarm.

Parameter	Description
Severity	Alarm severity. In versions earlier than MRS 3.x, the cluster alarm severity is as follows: ● Critical Indicates alarms reporting errors that affect cluster running, such as unavailable cluster services, node faults, data inconsistency between the active and standby GaussDB databases, and abnormal LdapServer data synchronization. You need to check the cluster status based on the alarms and rectify the faults in a timely manner. ● Major Indicates alarms reporting errors that affect some cluster functions, including process faults, periodic backup task failures, and abnormal key file permissions. Check the objects for which the alarms are generated based on the alarms and clear the alarms in a timely manner. ● Minor Indicates alarms reporting errors that do not affect major functions of the current cluster, including alarms indicating that the certificate file is about to expire, audit logs fail to be dumped, and the license file is about to expire. ● Warning Indicates an alarm of the lowest severity. It is used for information display or prompt and indicates that an event occurs in the scenarios when you stop a service, delete a service, stop an instance, delete an instance, delete a node, restart a service, restart an instance, perform an active/standby switchover for MRS Manager, scale in a host, or restore an instance. Additionally, this type of alarms also occurs when an instance is faulty, a job executed successfully, or a job failed to be executed. In MRS 3.x or later, the alarm severity of a cluster is as follows: ● Critical Indicates alarms reporting errors that affect cluster running, such as unavailable cluster services, node faults, data inconsistency between the active and standby GaussDB databases, and abnormal LdapServer data synchronization. You need to check the cluster status based on the alarms and rectify the faults in a timely manner. ● Major Indicates alarms reporting errors that affect some cluster functions, including process faults, periodic backup task failures, and abnormal key file permissions. Check the objects for which the alarms are generated based on the alarms and clear the alarms in a timely manner. ● Minor Indicates alarms reporting errors that do not affect major functions of the current cluster, including alarms indicating

Parameter	Description
	that the certificate file is about to expire, audit logs fail to be dumped, and the license file is about to expire. ● Suggestion Indicates an alarm of the lowest severity. It is used for information display or prompt and indicates that an event occurs in the scenarios when you stop a service, delete a service, stop an instance, delete an instance, delete a node, restart a service, restart an instance, perform an active/standby switchover for MRS Manager, scale in a host, or restore an instance. Additionally, this type of alarms also occurs when an instance is faulty, a job executed successfully, or a job failed to be executed.
Generated	Time when the alarm is generated.
Location	Details about the alarm.
Operation	If the alarm can be manually cleared, click **Clear Alarm**.

Table 4-48 Button description

Button	Description
Refresh every 30s ▾	Select an interval for refreshing the alarm list from the drop-down list. ● Refresh every 30s ● Refresh every 60s ● Stop refreshing
All ▾	Select an alarm severity from the drop-down list box to filter alarms. For versions earlier than MRS 3.x, the following alarms can be filtered: All, Critical, Major, Minor, and Warning. (For MRS 3.x or later) You can filter the following alarms: All, Critical, Major, Minor, and Warning.
↻	Click ↻ and manually refresh the alarm list.
Advanced Search	Click **Advanced Search**. In the displayed alarm search area, set search criteria and click **Search** to view the information about specified alarms. You can click **Reset** to clear the search criteria.

4.7.2 Viewing the Event List

The event list displays information about all events in a cluster, such as service restart and service termination.

Events are listed in chronological order by default in the event list, with the most recent events displayed at the top.

Table 4-49 describes various fields for an event.

Table 4-49 Event description

Parameter	Description
Event ID	Specifies the ID of an event.
Event Severity	Specifies the event severity. In versions earlier than MRS 3.x, the cluster event level is as follows: ● Critical ● Major ● Minor ● Suggestion In MRS 3.x or later, the event level of a cluster is as follows: ● Critical ● Major ● Minor ● Suggestion
Event Name	Name of the generated event.
Generated	Time when the event is generated.
Location	Specifies the detailed information for locating the event,

Table 4-50 Icon description

Icon	Description
Refresh every 30s ▾	Select an interval for refreshing the event list from the drop-down list. ● Refresh every 30s ● Refresh every 60s ● Stop refreshing
↻	Click ↻ to manually refresh the event list.
Advanced Search	Click **Advanced Search**. In the displayed event search area, set search criteria and click **Search** to view the information about specified events. Click **Reset** to clear the search criteria.

Exporting events

Step 1 Choose **Clusters > Active Clusters** and click a cluster name to go to the cluster details page.

Step 2 Click **Alarm Management** > **Events**.

Step 3 Click **Export All**.

Step 4 In the displayed dialog box, select the type and click **OK**.

----End

Common Events

Table 4-51 Common events

Event ID	Event Name
12019	Stop Service
12020	Delete Service
12021	Stop RoleInstance
12022	Delete RoleInstance
12023	Delete Node
12024	Restart Service
12025	Restart RoleInstance
12026	Manager Switchover
12065	Process Restart
12070	Job Running Succeeded
12071	Job Running Failed
12072	Job killed
12086	Agent Restart
14005	NameNode Switchover
14028	HDFS DiskBalancer Task
14029	Active NameNode enters safe mode and generates new Fsimage
17001	Oozie Workflow Execution Failure
17002	Oozie Scheduled Job Execution Failure
18001	ResourceManager Switchover
18004	JobHistoryServer Switchover

Event ID	Event Name
19001	HMaster Failover
20003	Hue Failover
24002	Flume Channel Overflow
25001	LdapServer Failover
27000	DBServer Switchover
38003	Adjusts the topic data storage period
43014	Spark Data Skew
43015	Spark SQL Large Query Results
43016	Spark SQL Execution Timeout
43024	Start JDBCServer
43025	Stop JDBCServer
43026	ZooKeeper Connection Succeeded
43027	Zookeeper Connection Failed

4.7.3 Viewing and Manually Clearing an Alarm

Scenario

You can view and clear alarms on MRS.

Generally, the system automatically clears an alarm when the fault is rectified. If the fault has been rectified and the alarm cannot be automatically cleared, you can manually clear the alarm.

You can view the latest 100,000 alarms (including uncleared, manually cleared, and automatically cleared alarms) on MRS. If the number of cleared alarms exceeds 100,000 and is about to reach 110,000, the system automatically dumps the earliest 10,000 cleared alarms to the dump path.

3. In versions earlier than x, the value is the same as that of ${BIGDATA_HOME}/OMSV100R001C00x8664/workspace/data for the active management node.

(For 3.x and later versions) The path is **${BIGDATA_HOME}/om-server/OMS/workspace/data** of the active management node.

A directory is automatically generated when alarms are dumped for the first time.

NOTE

Set an automatic refresh interval or click ↻ for an immediate refresh.

The following refresh interval options are supported:

- Refresh every 30 seconds
- Refresh every 60 seconds
- Stop refreshing

Procedure

Step 1 Choose **Clusters > Active Clusters** and click a cluster name to go to the cluster details page.

Step 2 Click **Alarms** and view the alarm information in the alarm list.

- By default, the alarm list page displays the latest 10 alarms.
- By default, data is sorted in descending order based on the generation time. For MRS 3.x or earlier, you can click the alarm ID, severity, and generation time to modify the sorting mode. For clusters of MRS 3.x or later, you can click the severity and generation time to modify the sorting mode.
- You can filter all alarms of the same severity. The results include cleared and uncleared alarms.
- For clusters of MRS 3.x and earlier versions, you can click 🔴, 🟠, 🟡 or 🔵 in the upper right corner of the page to quickly filter **Critical**, **Major**, **Minor**, or **Suggestion** alarms that are uncleared.
- For clusters of MRS 3.x or later: You can click 🔴, 🟠, 🟡 or 🔵 in the upper right corner of the page to quickly filter uncleared **Critical**, **Major**, **Minor** or **Warning** alarms.

Step 3 Click **Advanced Search**. In the displayed alarm search area, set search criteria and click **Search** to view the information about specified alarms. You can click **Reset** to clear the search criteria.

NOTE

The start time and end time are specified in **Time Range**. You can search for alarms generated within the time range.

Handle the alarm by referring to **Alarm Reference**. If the alarms in some scenarios are generated due to other cloud services that MRS depends on, you need to contact maintenance personnel of the corresponding cloud services.

Step 4 If the alarm needs to be manually cleared after errors are rectified, click **Clear Alarm**.

NOTE

If multiple alarms have been handled, you can select one or more alarms to be cleared and click **Clear Alarm** to clear the alarms in batches. A maximum of 300 alarms can be cleared in each batch.

----End

Exporting Alarms

Step 1 Choose **Clusters > Active Clusters** and click a cluster name to go to the cluster details page.

Step 2 Click **Alarm Management** > **Alarms**.

Step 3 Click **Export All**.

Step 4 In the displayed dialog box, select the type and click **OK**.

----End

4.8 Patch Management

4.8.1 Patch Operation Guide for Versions Earlier than MRS 1.7.0

If you obtain patch information from the following sources, upgrade the patch according to actual requirements.

- You obtain information about the patch released by MRS from a message pushed by the message center service.
- You obtain information about the patch by accessing the cluster and viewing patch information.

Preparing for Patch Installation

- Follow instructions in **Performing a Health Check** to check cluster status. If the cluster health status is normal, install a patch.
- You have uploaded the cluster patch package to the server. For details, see **Uploading a Patch Package**.
- You need to confirm the target patch to be installed according to the patch information in the patch content.

Uploading a Patch Package

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 Choose **System > Manage Patch**. The **Manage Patch** page is displayed.

Step 3 Click **Upload Patch** and set the following parameters.

- **Patch File Path**: folder created in the OBS file system where the patch package is stored, for example, **MRS_1.6.2/MRS_1_6_2_11.tar.gz**
- **Parallel File System Name**: name of the OBS file system that stores patch packages, for example, **mrs_patch**.

NOTE

You can obtain the file system name and patch file path on the **Patch Information** tab page. The value of the **Patch Path** is in the following format: *[File system name]/[Patch file path]*.

- **AK**: For details, see .
- **SK**: For details, see .

Step 4 Click **OK** to upload the patch.

----End

Installing a Patch

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 Choose **System** > **Manage Patch**. The **Manage Patch** page is displayed.

Step 3 In the **Operation** column, click **Install**.

Step 4 In the displayed dialog box, click **OK** to install the patch.

Step 5 After the patch is installed, you can view the installation status in the progress bar. If the installation fails, contact the MRS cluster administrator.

NOTE

For the isolated host nodes in the cluster, follow instructions in **Restoring Patches for the Isolated Hosts** to restore the patch.

----End

Uninstalling a Patch

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 Choose **System** > **Manage Patch**. The **Manage Patch** page is displayed.

Step 3 In the **Operation** column, click **Uninstall**.

NOTE

For the isolated host nodes in the cluster, follow instructions in **Restoring Patches for the Isolated Hosts** to restore the patch.

----End

4.8.2 Patch Operation Guide for

If you obtain patch information from the following sources, upgrade the patch according to actual requirements.

- You obtain information about the patch released by MRS from a message pushed by the message center service.
- You obtain information about the patch by accessing the cluster and viewing patch information.

Preparing for Patch Installation

- Follow instructions in **Performing a Health Check** to check cluster status. If the cluster health status is normal, install a patch.

- You need to confirm the target patch to be installed according to the patch information in the patch content.

Installing a Patch

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters** > **Active Clusters** and click the name of the cluster to be queried to enter the page displaying the cluster's basic information.

Step 3 On the **Patches** tab page, click **Install** in the **Operation** column to install the target patch.

NOTE

- For the isolated host nodes in the cluster, follow instructions in **Restoring Patches for the Isolated Hosts** to restore the patch.

----End

Uninstalling a Patch

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters** > **Active Clusters** and click the name of the cluster to be queried to enter the page displaying the cluster's basic information.

Step 3 On the **Patches** page, click **Uninstall** in the **Operation** column to uninstall the target patch.

NOTE

- For the isolated host nodes in the cluster, follow instructions in **Restoring Patches for the Isolated Hosts** to restore the patch.

----End

4.8.3 Rolling Patches

The rolling patch function indicates that patches are installed or uninstalled for one or more services in a cluster by performing a rolling service restart (restarting services or instances in batches), without interrupting the services or within a minimized service interruption interval. Services in a cluster are divided into the following three types based on whether they support rolling patch:

- Services supporting rolling patch installation or uninstallation: All businesses or part of them (varying depending on different services) of the services are not interrupted during patch installation or uninstallation.
- Services not supporting rolling patch installation or uninstallation: Businesses of the services are interrupted during patch installation or uninstallation.
- Services with some roles supporting rolling patch installation or uninstallation: Some businesses of the services are not interrupted during patch installation or uninstallation.

NOTE

In **MRS 3.x**, you cannot perform operations in this section on the management console.

Table 4-52 provides services and instances that support or do not support rolling restart in the MRS cluster.

Table 4-52 Services and instances that support or do not support rolling restart

Service	Instance	Whether to Support Rolling Restart
HDFS	NameNode	Yes
	Zkfc	
	JournalNode	
	HttpFS	
	DataNode	
Yarn	ResourceManager	Yes
	NodeManager	
Hive	MetaStore	Yes
	WebHCat	
	HiveServer	
MapReduce	JobHistoryServer	Yes
HBase	HMaster	Yes
	RegionServer	
	ThriftServer	
	RESTServer	
Spark	JobHistory	Yes
	JDBCServer	
	SparkResource	No
Hue	Hue	No
Tez	TezUI	No
Loader	Sqoop	No
Zookeeper	Quorumpeer	Yes
Kafka	Broker	Yes
	MirrorMaker	No
Flume	Flume	Yes
	MonitorServer	
Storm	Nimbus	Yes

Service	Instance	Whether to Support Rolling Restart
	UI	
	Supervisor	
	LogViewer	

Installing a Patch

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters** > **Active Clusters** and click the name of the cluster to be queried to enter the page displaying the cluster's basic information.

Step 3 On the **Patches** page, click **Install** in the **Operation** column.

Step 4 On the **Warning** page, enable or disable **Rolling Patch**.

NOTE

- Enabling the rolling patch installation function: Services are not stopped before patch installation, and rolling service restart is performed after the patch installation. This minimizes the impact on cluster services but takes more time than common patch installation.
- Disabling the rolling patch uninstallation function: All services are stopped before patch uninstallation, and all services are restarted after the patch uninstallation. This temporarily interrupts the cluster and the services but takes less time than rolling patch uninstallation.
- The rolling patch installation function is not available in clusters with less than two Master nodes and three Core nodes.

Step 5 Click **Yes** to install the target patch.

Step 6 View the patch installation progress.

1. Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.
2. Choose **System** > **Manage Patch**. On the **Manage Patch** page, you can view the patch installation progress.

NOTE

For the isolated host nodes in the cluster, follow instructions in **Restoring Patches for the Isolated Hosts** to restore the patch.

----End

Uninstalling a Patch

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters** > **Active Clusters** and click the name of the cluster to be queried to enter the page displaying the cluster's basic information.

Step 3 On the **Patches** page, click **Uninstall** in the **Operation** column.

Step 4 On the **Warning** page, enable or disable **Rolling Patch**.

NOTE

- Enabling the rolling patch uninstallation function: Services are not stopped before patch uninstallation, and rolling service restart is performed after the patch uninstallation. This minimizes the impact on cluster services but takes more time than common patch uninstallation.
- Disabling the rolling patch uninstallation function: All services are stopped before patch uninstallation, and all services are restarted after the patch uninstallation. This temporarily interrupts the cluster and the services but takes less time than rolling patch uninstallation.
- Only patches that are installed in rolling mode can be uninstalled in the same mode.

Step 5 Click **Yes** to uninstall the target patch.

Step 6 View the patch uninstallation progress.

1. Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.
2. Choose **System** > **Manage Patch**. On the **Manage Patch** page, you can view the patch uninstallation progress.

NOTE

For the isolated host nodes in the cluster, follow instructions in **Restoring Patches for the Isolated Hosts** to restore the patch.

----End

4.8.4 Restoring Patches for the Isolated Hosts

If some hosts are isolated in a cluster, perform the following operations to restore patches for these isolated hosts after patch installation on other hosts in the cluster. After patch restoration, versions of the isolated host nodes are consistent with those are not isolated.

NOTE

In **MRS 3.x**, you cannot perform operations in this section on the management console.

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 Choose **System > Manage Patch**. The **Manage Patch** page is displayed.

Step 3 In the **Operation** column, click **View Details**.

Step 4 On the patch details page, select host nodes whose **Status** is **Isolated**.

Step 5 Click **Select and Restore** to restore the isolated host nodes.

----End

4.9 Tenant Management

4.9.1 Before You Start

This section describes how to manage tenants on the MRS console.

Tenant management operations on the console apply only to clusters of versions earlier than MRS 3.x.

Tenant management operations on FusionInsight Manager apply to all versions. For MRS 3.x and later versions, see **Overview**. For versions earlier than MRS 3.x, see **Overview**.

4.9.2 Overview

Definition

An MRS cluster provides various resources and services for multiple organizations, departments, or applications to share. The cluster provides tenants as a logical entity to use these resources and services. A mode involving different tenants is called multi-tenant mode. Currently, only the analysis cluster supports tenant management.

Principles

The MRS cluster provides the multi-tenant function. It supports a layered tenant model and allows dynamic adding or deleting of tenants to isolate resources. It dynamically manages and configures tenants' computing and storage resources.

The computing resources indicate tenants' Yarn task queue resources. The task queue quota can be modified, and the task queue usage status and statistics can be viewed.

The storage resources can be stored on HDFS. You can add and delete the HDFS storage directories of tenants, and set the quotas of file quantity and the storage space of the directories.

Tenants can create and manage tenants in a cluster based on service requirements.

- Roles, computing resources, and storage resources are automatically created when tenants are created. By default, all permissions of the new computing resources and storage resources are allocated to a tenant's roles.
- Permissions to view the current tenant's resources, add a subtenant, and manage the subtenant's resources are granted to the tenant's roles by default.
- After you have modified the tenant's computing or storage resources, permissions of the tenant's roles are automatically updated.

MRS supports a maximum of 512 tenants. The default tenants created by the system include **default**. Tenants that are in the topmost layer with the default tenant are called level-1 tenants.

Resource Pools

Yarn task queues support only the label-based scheduling policy. This policy enables Yarn task queues to associate NodeManagers that have specific node labels. In this way, Yarn tasks run on specified nodes so that tasks are scheduled and certain hardware resources are utilized. For example, Yarn tasks requiring a large memory capacity can run on nodes with a large memory capacity by means of label association, preventing poor service performance.

In an MRS cluster, the tenant logically divides Yarn cluster nodes to combine multiple NodeManagers into a resource pool. Yarn task queues can be associated with specified resource pools by configuring queue capacity policies, ensuring efficient and independent resource utilization in the resource pools.

MRS supports a maximum of 50 resource pools. By default, the system contains a **default** resource pool.

4.9.3 Creating a Tenant

Scenario

You can create a tenant on MRS Manager to specify the resource usage.

Prerequisites

- A tenant name has been planned. The name must not be the same as that of a role or Yarn queue that exists in the current cluster.
- If a tenant requires storage resources, a storage directory has been planned based on service requirements, and the planned directory does not exist under the HDFS directory.
- The resources that can be allocated to the current tenant have been planned and the sum of the resource percentages of direct sub-tenants under the parent tenant at every level does not exceed 100%.
- You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS details page, click **Tenants**.

NOTE

For MRS 3.x or later, see **Overview**.

Step 2 Click **Create Tenant**. On the page that is displayed, configure tenant properties.

Table 4-53 Tenant parameters

Parameter	Description
Name	Specifies the name of the current tenant. The value consists of 3 to 50 characters, and can contain letters, digits, and underscores (_).
Tenant Type	The options include **Leaf** and **Non-leaf**. If **Leaf** is selected, the current tenant is a leaf tenant and no sub-tenant can be added. If **Non-leaf** is selected, sub-tenants can be added to the current tenant.

Parameter	Description
Dynamic Resource	Specifies the dynamic computing resources for the current tenant. The system automatically creates a task queue named after the tenant name in Yarn. When dynamic resources are not **Yarn**, the system does not automatically create a task queue.
Default Resource Pool Capacity (%)	Specifies the percentage of the computing resources used by the current tenant in the **default** resource pool.
Default Resource Pool Max. Capacity (%)	Specifies the maximum percentage of the computing resources used by the current tenant in the **default** resource pool.
Storage Resource	Specifies storage resources for the current tenant. The system automatically creates a file folder named after the tenant name in the **/tenant** directory. When a tenant is created for the first time, the system automatically creates the **/tenant** directory in the HDFS root directory. If storage resources are not **HDFS**, the system does not create a storage directory under the root directory of HDFS.
Space Quota (MB)	Specifies the quota for HDFS storage space used by the current tenant. The value ranges from **1** to **8796093022208**. The unit is MB. This parameter indicates the maximum HDFS storage space that can be used by a tenant, but does not indicate the actual space used. If the value is greater than the size of the HDFS physical disk, the maximum space available is the full space of the HDFS physical disk. **NOTE** To ensure data reliability, one backup is automatically generated for each file saved in HDFS, that is, two copies are generated in total. The HDFS storage space indicates the total disk space occupied by all these copies. For example, if the value is set to **500**, the actual space for storing files is about 250 MB (500/2 = 250).
Storage Path	Specifies the tenant's HDFS storage directory. The system automatically creates a file folder named after the tenant name in the **/tenant** directory by default. For example, the default HDFS storage directory for **ta1** is **tenant/ta1**. When a tenant is created for the first time, the system automatically creates the **/tenant** directory in the HDFS root directory. The storage path is customizable.

Parameter	Description
Service	Specifies other service resources associated with the current tenant. HBase is supported. To configure this parameter, click **Associate Services**. In the dialog box that is displayed, set **Service** to **HBase**. If **Association Mode** is set to **Exclusive**, service resources are occupied exclusively. If **share** is selected, service resources are shared.
Description	Specifies the description of the current tenant.

Step 3 Click **OK** to save the settings.

It takes a few minutes to save the settings. If the **Tenant created successfully** is displayed in the upper-right corner, the tenant is added successfully. The tenant is created successfully.

NOTE

- Roles, computing resources, and storage resources are automatically created when tenants are created.
- The new role has permissions on the computing and storage resources. The role and its permissions are controlled by the system automatically and cannot be controlled manually under **Manage Role**.
- If you want to use the tenant, create a system user and assign the Manager_tenant role and the role corresponding to the tenant to the user. For details, see **Creating a User**.

----End

Related Tasks

View an added tenant.

Step 1 On the MRS details page, click **Tenants**.

Step 2 In the tenant list on the left, click the name of the added tenant.

The **Summary** tab is displayed on the right by default.

Step 3 View **Basic Information**, **Resource Quota**, and **Statistics** of the tenant.

If HDFS is in the **Stopped** state, **Available** and **Used** of **Space** in **Resource Quota** are **unknown.**

----End

4.9.4 Creating a Sub-tenant

Scenario

You can create a sub-tenant on MRS if the resources of the current tenant need to be further allocated.

Prerequisites

- A parent tenant has been added.
- A tenant name has been planned. The name must not be the same as that of a role or Yarn queue that exists in the current cluster.
- If a sub-tenant requires storage resources, a storage directory has been planned based on service requirements, and the planned directory does not exist under the storage directory of the parent tenant.
- The resources that can be allocated to the current tenant have been planned and the sum of the resource percentages of direct sub-tenants under the parent tenant at every level does not exceed 100%.
- You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS details page, click **Tenants**.

NOTE

For MRS 3.x or later, see **Overview**.

Step 2 In the tenant list on the left, move the cursor to the tenant node to which a sub-tenant is to be added. Click **Create sub-tenant**. On the displayed page, configure the sub-tenant attributes according to the following table:

Table 4-54 Sub-tenant parameters

Parameter	Description
Parent tenant	Specifies the name of the parent tenant.
Name	Specifies the name of the current tenant. The value consists of 3 to 20 characters, and can contain letters, digits, and underscores (_).
Tenant Type	The options include **Leaf** and **Non-leaf**. If **Leaf** is selected, the current tenant is a leaf tenant and no sub-tenant can be added. If **Non-leaf** is selected, sub-tenants can be added to the current tenant.
Dynamic Resource	Specifies the dynamic computing resources for the current tenant. The system automatically creates a task queue named after the sub-tenant name in the Yarn parent queue. When dynamic resources are not **Yarn**, the system does not automatically create a task queue. If the parent tenant does not have dynamic resources, the sub-tenant cannot use dynamic resources.
Default Resource Pool Capacity (%)	Specifies the percentage of the resources used by the current tenant. The base value is the total resources of the parent tenant.

Parameter	Description
Default Resource Pool Max. Capacity (%)	Specifies the maximum percentage of the computing resources used by the current tenant. The base value is the total resources of the parent tenant.
Storage Resource	Specifies storage resources for the current tenant. The system automatically creates a file in the HDFS parent tenant directory. The file is named the same as the name of the sub-tenant. If storage resources are not **HDFS**, the system does not create a storage directory under the root directory of HDFS. If the parent tenant does not have storage resources, the sub-tenant cannot use storage resources.
Space Quota (MB)	Specifies the quota for HDFS storage space used by the current tenant. The minimum value is 1, and the maximum value is the total storage quota of the parent tenant. The unit is MB. This parameter indicates the maximum HDFS storage space that can be used by a tenant, but does not indicate the actual space used. If the value is greater than the size of the HDFS physical disk, the maximum space available is the full space of the HDFS physical disk. If the quota is greater than the quota of the parent tenant, the actual storage capacity is subject to the quota of the parent tenant. **NOTE** To ensure data reliability, one backup is automatically generated for each file saved in HDFS, that is, two copies are generated in total. The HDFS storage space indicates the total disk space occupied by all these copies. For example, if the value is set to **500**, the actual space for storing files is about 250 MB (500/2 = 250).
Storage Path	Specifies the tenant's HDFS storage directory. The system automatically creates a file folder named after the sub-tenant name in the directory of the parent tenant by default. For example, if the sub-tenant is **ta1s** and the parent directory is **tenant/ta1**, the system sets this parameter for the sub-tenant to **tenant/ta1/ta1s**. The storage path is customizable in the parent directory. The parent directory for the storage path must be the storage directory of the parent tenant.
Service	Specifies other service resources associated with the current tenant. HBase is supported. To configure this parameter, click **Associate Services**. In the dialog box that is displayed, set **Service** to **HBase**. If **Association Mode** is set to **Exclusive**, service resources are occupied exclusively. If **share** is selected, service resources are shared.
Description	Specifies the description of the current tenant.

Step 3 Click **OK** to save the settings.

It takes a few minutes to save the settings. If the **Tenant created successfully** is displayed in the upper-right corner, the tenant is added successfully. The tenant is created successfully.

NOTE

- Roles, computing resources, and storage resources are automatically created when tenants are created.
- The new role has permissions on the computing and storage resources. The role and its permissions are controlled by the system automatically and cannot be controlled manually under **Manage Role**.
- When using this tenant, create a system user and assign the user a related tenant role. For details, see **Creating a User**.

----End

4.9.5 Deleting a Tenant

Scenario

You can delete a tenant that is not required on MRS.

Prerequisites

- A tenant has been added.
- You have checked whether the tenant to be deleted has sub-tenants. If the tenant has sub-tenants, delete them; otherwise, you cannot delete the tenant.
- The role of the tenant to be deleted cannot be associated with any user or user group. For details about how to cancel the binding between a role and a user, see **Modifying User Information**.
- You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS details page, click **Tenants**.

NOTE

For MRS 3.x or later, see **Overview**.

Step 2 In the tenant list on the left, move the cursor to the tenant node to be deleted and click **Delete**.

The **Delete Tenant** dialog box is displayed. If you want to save the tenant data, select **Reserve the data of this tenant**. Otherwise, the tenant's storage space will be deleted.

Step 3 Click **OK**.

It takes a few minutes to save the configuration. After the tenant is deleted successfully, the role and storage space of the tenant are also deleted.

NOTE

- After the tenant is deleted, the task queue of the tenant still exists in Yarn.
- If you choose not to reserve data when deleting the parent tenant, data of sub-tenants is also deleted if the sub-tenants use storage resources.

----End

4.9.6 Managing a Tenant Directory

Scenario

You can manage the HDFS storage directory used by a specific tenant on MRS. The management operations include adding a tenant directory, modifying the directory file quota, modifying the storage space, and deleting a directory.

Prerequisites

- A tenant associated with HDFS storage resources has been added.
- You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

- View a tenant directory.
  a. On the MRS details page, click **Tenants**.

     NOTE

     For MRS 3.x or later, see **Overview**.

  b. In the tenant list on the left, click the target tenant.
  c. Click the **Resources** tab.
  d. View the **HDFS Storage** table.
     - The **Maximum Number of Files/Directories** column indicates the quotas for the file and directory quantity of the tenant directory.
     - The **Space Quota** column indicates storage space size of tenant directories.
- Add a tenant directory.
  a. On the MRS details page, click **Tenants**.

     NOTE

     For MRS 3.x or later, see **Overview**.

  b. In the tenant list on the left, click the tenant whose HDFS storage directory needs to be added.
  c. Click the **Resources** tab.
  d. In the **HDFS Storage** table, click **Create Directory**.
     - Set **Path** to a tenant directory path.

NOTE

- If the current tenant is not a sub-tenant, the new path is created in the HDFS root directory.
- If the current tenant is a sub-tenant, the new path is created in the specified directory.

A complete HDFS storage directory can contain a maximum of 1,023 characters. An HDFS directory name contains digits, letters, spaces, and underscores (_). The name cannot start or end with a space.

- Set **Maximum Number of Files/Directories** to the quotas of file and directory quantity.

  Maximum Number of Files/Directories is optional. Its value ranges from **1** to **9223372036854775806**.

- Set **Storage Space Quota** to the storage space size of the tenant directory.

  The value of **Storage Space Quota** ranges from **1** to **8796093022208**.

  NOTE

  To ensure data reliability, one backup is automatically generated for each file saved in HDFS, that is, two copies are generated in total. The HDFS storage space indicates the total disk space occupied by all these copies. For example, if the value of **Storage Space Quota** is set to **500**, the actual space for storing files is about 250 MB (500/2 = 250).

e. Click **OK**. The system creates tenant directories in the HDFS root directory.

- Modify a tenant directory.

  a. On the MRS details page, click **Tenants**.

  NOTE

  For MRS 3.x or later, see **Overview**.

  b. In the tenant list on the left, click the tenant whose HDFS storage directory needs to be modified.

  c. Click the **Resources** tab.

  d. In the **HDFS Storage** table, click **Modify** in the **Operation** column of the specified tenant directory.

  - Set **Maximum Number of Files/Directories** to the quotas of file and directory quantity.

    Maximum Number of Files/Directories is optional. Its value ranges from **1** to **9223372036854775806**.

  - Set **Storage Space Quota** to the storage space size of the tenant directory.

    The value of **Storage Space Quota** ranges from **1** to **8796093022208**.

NOTE

To ensure data reliability, one backup is automatically generated for each file saved in HDFS, that is, two copies are generated in total. The HDFS storage space indicates the total disk space occupied by all these copies. For example, if the value of **Storage Space Quota** is set to **500**, the actual space for storing files is about 250 MB (500/2 = 250).

   e. Click **OK**.

- Delete a tenant directory.
   a. On the MRS details page, click **Tenants**.

      NOTE

      For MRS 3.x or later, see **Overview**.

   b. In the tenant list on the left, click the tenant whose HDFS storage directory needs to be deleted.
   c. Click the **Resources** tab.
   d. In the **HDFS Storage** table, click **Delete** in the **Operation** column of the specified tenant directory.

      The default HDFS storage directory set during tenant creation cannot be deleted. Only the newly added HDFS storage directory can be deleted.
   e. Click **OK**. The tenant directory is deleted.

4.9.7 Restoring Tenant Data

Scenario

Tenant data is stored on Manager and in cluster components by default. When components are restored from faults or reinstalled, some tenant configuration data may be abnormal. In this case, you can manually restore the tenant data.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS details page, click **Tenants**.

NOTE

For MRS 3.x or later, see **Overview**.

Step 2 In the tenant list on the left, click a tenant node.

Step 3 Check the status of the tenant data.

1. In **Summary**, check the color of the circle on the left of **Basic Information**. Green indicates that the tenant is available and gray indicates that the tenant is unavailable.
2. Click **Resources** and check the circle on the left of **Yarn** or **HDFS Storage**. Green indicates that the resource is available, and gray indicates that the resource is unavailable.

3. Click **Service Association** and check the **Status** column of the associated service table. **Good** indicates that the component can provide services for the associated tenant. **Bad** indicates that the component cannot provide services for the tenant.
4. If any check result is abnormal, go to **Step 4** to restore tenant data.

Step 4 Click **Restore Tenant Data**.

Step 5 In the **Restore Tenant Data** window, select one or more components whose data needs to be restored. Click **OK**. The system automatically restores the tenant data.

----End

4.9.8 Creating a Resource Pool

Scenario

In an MRS cluster, users can logically divide Yarn cluster nodes to combine multiple NodeManagers into a Yarn resource pool. Each NodeManager belongs to one resource pool only. The system contains a **default** resource pool by default. All NodeManagers that are not added to customized resource pools belong to this resource pool.

You can create a customized resource pool on MRS and add hosts that have not been added to other customized resource pools to it.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS details page, click **Tenants**.

NOTE

For MRS 3.x or later, see **Overview**.

Step 2 Click the **Resource Pools** tab.

Step 3 Click **Create Resource Pool**.

Step 4 In **Create Resource Pool**, set the properties of the resource pool.

- **Name**: Enter a name for the resource pool. The name of the newly created resource pool cannot be **default**.

  The name consists of 1 to 20 characters and can contain digits, letters, and underscores (_) but cannot start with an underscore (_).
- **Available Hosts**: In the host list on the left, select a specified host name and add it to the resource pool. Only hosts in the cluster can be selected. The host list of a resource pool can be left blank.

Step 5 Click **OK**.

Step 6 After a resource pool is created, users can view the **Name**, **Members**, **Type**, **vCore** and **Memory** in the resource pool list. Hosts that are added to the customized resource pool are no longer members of the **default** resource pool.

----End

4.9.9 Modifying a Resource Pool

Scenario

You can modify members of an existing resource pool on MRS.

Prerequisites

You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS details page, click **Tenants**.

NOTE

For MRS 3.x or later, see **Overview**.

Step 2 Click the **Resource Pools** tab.

Step 3 Locate the row that contains the specified resource pool, and click **Modify** in the **Operation** column.

Step 4 In **Modify Resource Pool**, modify **Added Hosts**.

- Adding a host: In the host list on the left, select the specified host name and add it to the resource pool.
- Deleting a host: In the host list on the right, click ✕ next to a host to remove the host from the resource pool. The host list of a resource pool can be left blank.

Step 5 Click **OK**.

----End

4.9.10 Deleting a Resource Pool

Scenario

You can delete an existing resource pool on MRS.

Prerequisites

- Any queue in a cluster cannot use the resource pool to be deleted as the default resource pool. Before deleting the resource pool, cancel the default resource pool. For details, see **Configuring a Queue**.
- Resource distribution policies of all queues have been cleared from the resource pool being deleted. For details, see **Clearing Configuration of a Queue**.

- You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS details page, click **Tenant**.

NOTE

For MRS 3.x or later, see **Overview**.

Step 2 Click the **Resource Pools** tab.

Step 3 Locate the row that contains the specified resource pool, and click **Delete** in the **Operation** column.

In the displayed dialog box, click **OK**.

----End

4.9.11 Configuring a Queue

Scenario

You can modify the queue configuration of a specified tenant on MRS based on service requirements.

Prerequisites

- A tenant associated with Yarn and allocated dynamic resources has been added.
- You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS details page, click **Tenants**.

NOTE

For MRS 3.x or later, see **Overview**.

Step 2 Click the **Queue Configuration** tab.

Step 3 In the tenant queue table, click **Modify** in the **Operation** column of the specified tenant queue.

NOTE

- In the tenant list on the left of the **Tenant Management** tab, click the target tenant. In the window that is displayed, choose **Resource**. On the page that is displayed, click ✎ to open the queue modification page.
- A queue can be bound to only one non-default resource pool.

Versions earlier than MRS 3.x:

Table 4-55 Queue configuration parameters

Parameter	Description
Maximum Applications	Specifies the maximum number of applications. The value ranges from 1 to 2147483647.
Maximum AM Resource Percent	Specifies the maximum percentage of resources that can be used to run the ApplicationMaster in a cluster. The value ranges from 0 to 1.
Minimum User Limit Percent (%)	Specifies the minimum percentage of resources consumed by a user. The value ranges from 0 to 100.
User Limit Factor	Specifies the limit factor of the maximum user resource usage. The maximum user resource usage percentage can be obtained by multiplying the limit factor with the percentage of the tenant's actual resource usage in the cluster. The minimum value is **0**.
Status	Specifies the current status of a resource plan. The values are **Running** and **Stopped**.
Default Resource Pool (Default Node Label Expression)	Specifies the resource pool used by a queue. The default value is **default**. If you want to change the resource pool, configure the queue capacity first. For details, see **Configuring the Queue Capacity Policy of a Resource Pool**.

MRS 3.x or later:

Table 4-56 Queue configuration parameters

Parameter	Description
Max Master Shares (%)	Indicates the maximum percentage of resources occupied by all ApplicationMasters in the current queue.
Max Allocated vCores	Indicates the maximum number of cores that can be allocated to a single YARN container in the current queue. The default value is **-1**, indicating that the number of cores is not limited within the value range.
Max Allocated Memory (MB)	Indicates the maximum memory that can be allocated to a single Yarn container in the current queue. The default value is **-1**, indicating that the memory is not limited within the value range.

Parameter	Description
Max Running Apps	Maximum number of tasks that can be executed at the same time in the current queue. The default value is **-1**, indicating that the number is not limited within the value range (the meaning is the same if the value is empty). The value 0 indicates that the task cannot be executed. The value ranges from -1 to 2147483647.
Max Running Apps per User	Maximum number of tasks that can be executed by each user in the current queue at the same time. The default value is **-1**, indicating that the number is not limited within the value range. If the value is **0**, the task cannot be executed. The value ranges from -1 to 2147483647.
Max Pending Apps	Maximum number of tasks that can be suspended at the same time in the current queue. The default value is **-1**, indicating that the number is not limited within the value range (the meaning is the same if the value is empty). The value **0** indicates that tasks cannot be suspended. The value ranges from -1 to 2147483647.
Resource Allocation Rule	Indicates the rule for allocating resources to different tasks of a user. The rule can be FIFO or FAIR. If a user submits multiple tasks in the current queue and the rule is FIFO, the tasks are executed one by one in sequential order. If the rule is FAIR, resources are evenly allocated to all tasks.
Default Resource Label	Indicates that tasks are executed on a node with a specified resource label. **NOTE** If you need to use a new resource pool, change the default label to the new resource pool label.
Active	● **ACTIVE**: indicates that the current queue can receive and execute tasks. ● **INACTIVE**: indicates that the current queue can receive but cannot execute tasks. Tasks submitted to the queue are suspended.
Open	● **OPEN**: indicates that the current queue is opened. ● **CLOSED**: indicates that the current queue is closed. Tasks submitted to the queue are rejected.

----End

4.9.12 Configuring the Queue Capacity Policy of a Resource Pool

Scenario

After a resource pool is added, the capacity policies of available resources need to be configured for Yarn task queues. This ensures that tasks in the resource pool are running properly. Each queue can be configured with the queue capacity policy of only one resource pool. Users can view the queues in any resource pool and configure queue capacity policies. After the queue policies are configured, Yarn task queues and resource pools are associated.

You can configure queue policies on MRS.

Prerequisites

- A resource pool has been added.
- The task queues are not associated with other resource pools. By default, all queues are associated with the **default** resource pool.
- You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS details page, click **Tenants**.

NOTE

For MRS 3.x or later, see **Overview**.

Step 2 Click the **Resource Distribution Policies** tab.

Step 3 In **Resource Pools**, select a specified resource pool.

Available Resource Quota: indicates that all resources in each resource pool are available for queues by default.

Step 4 Locate the specified queue in the **Resource Allocation** table, and click **Modify** in the **Operation** column.

Step 5 In **Modify Resource Allocation**, configure the resource capacity policy of the task queue in the resource pool.

- **Capacity (%)**: specifies the percentage of the current tenant's computing resource usage.
- **Maximum Capacity (%)**: specifies the percentage of the current tenant's maximum computing resource usage.

Step 6 Click **OK** to save the settings.

----End

4.9.13 Clearing Configuration of a Queue

Scenario

Users can clear the configuration of a queue on MRS Manager when the queue does not need resources from a resource pool or if a resource pool needs to be disassociated from the queue. Clearing queue configurations means that the resource capacity policy of the queue is canceled.

Prerequisites

- If a queue is to be unbound from a resource pool, this resource pool cannot serve as the default resource pool of the queue. Therefore, you must first change the default resource pool of the queue to another one. For details, see **Configuring a Queue**.
- You have synchronized IAM users. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Procedure

Step 1 On the MRS details page, click **Tenants**.

NOTE

For MRS 3.x or later, see **Overview**.

Step 2 Click the **Resource Distribution Policies** tab.

Step 3 In **Resource Pools**, select a specified resource pool.

Step 4 Locate the specified queue in the **Resource Allocation** table, and click **Clear** in the **Operation** column

In the **Clear Queue Configuration** dialog box, click **OK** to clear the queue configuration in the current resource pool.

NOTE

If no resource capacity policy is configured for a queue, the clearing function is unavailable for the queue by default.

----End

4.10 Bootstrap Actions

4.10.1 Introduction to Bootstrap Actions

Bootstrap actions indicate that you can run your scripts on a specified cluster node before or after starting big data components. You can run bootstrap actions to install additional third-party software, modify the cluster running environment, and perform other customizations.

If you choose to run bootstrap actions when scaling out a cluster, the bootstrap actions will be run on the newly added nodes in the same way. If auto scaling is enabled in a cluster, you can add an automation script in addition to configuring a resource plan. Then the automation script executes the corresponding script on the nodes that are scaled out or in to implement custom operations.

For versions earlier than MRS 3.*x*, your scripts are executed as user **root**. You can run the **su - XXX** command in a script to switch to another user.

For MRS 3.*x* or later, your scripts are executed as user **omm**. You can run the **su - XXX** command in a script to switch to another user.

NOTE

Versions earlier than MRS 3.*x*: Bootstrap action scripts must be executed as user **root**. Otherwise, your cluster may become unavailable.

MRS 3.*x* or later: Bootstrap action scripts must be executed as user **omm**. Otherwise, your cluster may become unavailable.

MRS determines the result based on the return code after the execution of the bootstrap action script. If the return code is **0**, the script is executed successfully. If the return code is not **0**, the execution fails. If a bootstrap action script fails to be executed on a node, the corresponding boot script will fail to be executed. In this case, you can set **Action upon Failure** to choose whether to continue to execute the subsequent scripts. Example 1: If you set **Action upon Failure** to **Continue** for all scripts during cluster creation, all the scripts will be executed regardless of whether they are successfully executed, and the startup process will be complete. Example 2: If a script fails to be executed and **Action upon Failure** is set to **Stop**, subsequent scripts will not be executed and cluster creation or scale-out will fail.

You can add a maximum of 18 bootstrap actions, which will be executed before or after the cluster component is started in the order you specified. The bootstrap actions performed before or after the component startup must be completed within 60 minutes. Otherwise, the cluster creation or scale-out will fail.

4.10.2 Preparing the Bootstrap Action Script

Currently, bootstrap actions support Linux shell scripts only. Script files must end with **.sh**.

Uploading the Installation Packages and Files to an OBS File System

Before compiling a script, you need to upload all required installation packages, configuration packages, and relevant files to the OBS file system in the same region. Because networks of different regions are isolated from each other, MRS VMs cannot download OBS files from other regions.

Compiling a Script for Downloading Files from the OBS File System

You can specify the file to be downloaded from OBS in the script. If you upload files to a private file system, you need to run the **hadoop fs** command to download the files. The following example shows that the **obs://yourbucket/myfile.tar.gz** file will be downloaded to the local host and decompressed to the **/your-dir** directory.

```
#!/bin/bash
source /opt/Bigdata/client/bigdata_env;hadoop fs -D fs.obs.endpoint=<obs-endpoint> -D 
fs.obs.access.key=<your-ak> -D fs.obs.secret.key=<your-sk> -copyToLocal obs://yourbucket/
```

```
myfile.tar.gz ./
mkdir -p /<your-dir>
tar -zxvf myfile.tar.gz -C /<your-dir>
```

NOTE

- The default client installation path for MRS 3.x or later is /opt/Bigdata/client, and for versions earlier than MRS 3.x is /opt/client. Configure the path based on site requirements.
- The Hadoop client has been preinstalled on the MRS node. You can run the **hadoop fs** command to download or upload data from or to OBS.
- Obtain the obs-endpoint of each region. For details, see **Regions and Endpoints**.

Uploading the Script to the OBS File System

After script compilation, upload the script to the OBS file system in the same region. At the time you specify, each node in the cluster downloads the script from OBS and executes the script as user **root**.

4.10.3 View Execution Records

You can view the execution result of the bootstrap operation on the **Bootstrap Action** page.

Viewing the Execution Result

1. Log in to the MRS console.
2. In the left navigation pane, choose **Clusters** > **Active Clusters**. Click a cluster you want to query.

   The cluster details page is displayed.
3. On the cluster details page, click the **Bootstrap Action** tab. Information about the bootstrap actions added during cluster creation is displayed.

   NOTE

   - You select **Before initial component start** or **After initial component start** in the upper right corner to query information about the related bootstrap actions.
   - The last execution result is listed here. For a newly created cluster, the records of bootstrap actions executed during cluster creation are listed. If a cluster is expanded, the records of bootstrap actions executed on the newly added nodes are listed.

Viewing Execution Logs

If you want to view the run logs of a bootstrap action, set **Action upon Failure** to **Continue** when adding the bootstrap action. And then, log in to each node to view the run logs in the **/var/log/Bootstrap** directory. If you add bootstrap actions before and after component start, you can distinguish bootstrap action logs of the two phases based on the timestamps.

You are advised to print logs in detail in the script so that you can view the detailed run result. MRS redirects the standard output and error output of the script to the log directory of the bootstrap action.

4.10.4 Adding a Bootstrap Action

Add a bootstrap action.

This operation applies to MRS 3.*x* or earlier clusters.

Procedure

Step 1 Log in to the MRS management console.

Step 2 Choose **Clusters** > **Active Clusters** and click the name of your desired cluster.

Step 3 On the page that is displayed, click the **Bootstrap Actions** tab.

Step 4 Click **Add** and set parameters as prompted.

Table 4-57 Parameters

Parameter	Description
Name	Name of a bootstrap action script The value can contain only digits, letters, spaces, hyphens (-), and underscores (_) and must not start with a space. The value can contain 1 to 64 characters. **NOTE** A name must be unique in the same cluster. You can set the same name for different clusters.
Script Path	Script path. The value can be an OBS file system path or a local VM path. ● An OBS file system path must start with **s3a://** and end with **.sh**, for example, **s3a://mrs-samples/*xxx*.sh**. ● A local VM path must start with a slash (/) and end with **.sh**. **NOTE** A path must be unique in the same cluster, but can be the same for different clusters.
Parameter	Bootstrap action script parameters
Execution Node	Select a type of the node where the bootstrap action script is executed.
Executed	Select the time when the bootstrap action script is executed. ● Before initial component start ● After initial component start
Action upon Failure	Whether to continue to execute subsequent scripts and create a cluster after the script fails to be executed. **NOTE** You are advised to set this parameter to **Continue** in the debugging phase so that the cluster can continue to be installed and started no matter whether the bootstrap action is successful.

Step 5 Click **OK** to save the configuration.

Step 6 Click **Yes**.

----End

4.10.5 Modifying a Bootstrap Action

Scenario

Modify an existing bootstrap action on an MRS cluster.

Procedure

Step 1 Log in to the MRS management console.

Step 2 Choose **Clusters** > **Active Clusters** and click the name of your desired cluster.

Step 3 On the page that is displayed, click the **Bootstrap Actions** tab.

Step 4 In the list, select the item to be modified and click **Edit**.

Step 5 Modify the parameters as needed.

Step 6 Click **OK** to save the modification.

Step 7 Click **Yes**.

----End

4.10.6 Deleting a Bootstrap Action

Scenario

Delete a bootstrap action on an MRS cluster.

Procedure

Step 1 Log in to the MRS management console.

Step 2 Choose **Clusters** > **Active Clusters** and click the name of your desired cluster.

Step 3 On the page that is displayed, click the **Bootstrap Actions** tab.

Step 4 In the list, select the item to be deleted and click **Delete**.

Step 5 Click **OK**.

----End

5 Using an MRS Client

5.1 Installing a Client

5.1.1 Installing a Client (Version 3.x or Later)

Scenario

This section describes how to install clients of all services (excluding Flume) in an MRS cluster. For details about how to install the Flume client, see .

A client can be installed on a node inside or outside the cluster. This section uses the installation directory **//opt/client** as an example. Replace it with the actual one.

Prerequisites

- A Linux ECS has been prepared. For details about the supported OS of the ECS, see **Table 5-1**.

Table 5-1 Reference list

CPU Architecture	OS	Supported Version
x86 computing	Euler	EulerOS 2.5
	SUSE	SUSE Linux Enterprise Server 12 SP4 (SUSE 12.4)
	RedHat	Red Hat-7.5-x86_64 (Red Hat 7.5)
	CentOS	CentOS 7.6
Kunpeng computing (Arm)	Euler	EulerOS 2.8
	CentOS	CentOS 7.6

In addition, sufficient disk space is allocated for the ECS, for example, 40 GB.

- The ECS and the MRS cluster are in the same VPC.
- The security group of the ECS must be the same as that of the master node in the MRS cluster.
- The NTP service has been installed on the ECS OS and is running properly.

  If the NTP service is not installed, run the **yum install ntp -y** command to install it when the **yum** source is configured.
- A user can log in to the Linux ECS using the password (in SSH mode).

Installing a Client on a Node Inside a Cluster

1. Obtain the software package.

   Log in to FusionInsight Manager. For details, see **Accessing FusionInsight Manager (MRS 3.x or Later)**. Click the name of the cluster to be operated in the **Cluster** drop-down list.

   Choose **More** > **Download Client**. The **Download Cluster Client** dialog box is displayed.

   NOTE

   In the scenario where only one client is to be installed, choose **Cluster** > **Service** > *Service name* > **More** > **Download Client**. The **Download Client** dialog box is displayed.

2. Set the client type to **Complete Client**.

   Configuration Files Only is to download client configuration files in the following scenario: After a complete client is downloaded and installed and modify server configurations on Manager, developers need to update the configuration files during application development.

   The platform type can be set to **x86_64** or **aarch64**.

   - **x86_64**: indicates the client software package that can be deployed on the x86 servers.
   - **aarch64**: indicates the client software package that can be deployed on the TaiShan servers.

   NOTE

   The cluster supports two types of clients: **x86_64** and **aarch64**. The client type must match the architecture of the node for installing the client. Otherwise, client installation will fail.

3. Select **Save to Path** and click **OK** to generate the client file.

   The generated file is stored in the **/tmp/FusionInsight-Client** directory on the active management node by default. You can also store the client file in a directory on which user **omm** has the read, write, and execute permissions. Copy the software package to the file directory on the server where the client is to be installed as user **omm** or **root**.

   The name of the client software package is in the follow format: **FusionInsight_Cluster_***<Cluster ID>***_Services_Client.tar**. In this section, the cluster ID **1** is used as an example. Replace it with the actual cluster ID.

The following steps and sections use **FusionInsight_Cluster_1_Services_Client.tar** as an example.

NOTE

If you cannot obtain the permissions of user **root**, use user **omm**.

To install the client on another node in the cluster, run the following command to copy the client to the node where the client is to be installed:

scp -p /*tmp/FusionInsight-Client***/FusionInsight_Cluster_1_Services_Client.tar** *IP address of the node where the client is to be installed:/opt/Bigdata/client*

4. Log in to the server where the client software package is located as user **user_client**.
5. Decompress the software package.

   Go to the directory where the installation package is stored, such as **/tmp/FusionInsight-Client**. Run the following command to decompress the installation package to a local directory:

   tar -xvf FusionInsight_Cluster_1_Services_Client.tar
6. Verify the software package.

   Run the following command to verify the decompressed file and check whether the command output is consistent with the information in the **sha256** file.

   sha256sum -c FusionInsight_Cluster_1_Services_ClientConfig.tar.sha256

   ```
   FusionInsight_Cluster_1_Services_ClientConfig.tar: OK
   ```
7. Decompress the obtained installation file.

   tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar
8. Go to the directory where the installation package is stored, and run the following command to install the client to a specified directory (an absolute path), for example, **/opt/client**:

   cd /tmp/FusionInsight-Client/FusionInsight_Cluster_1_Services_ClientConfig

   Run the **./install.sh /opt/client** command to install the client. The client is successfully installed if information similar to the following is displayed:

   ```
   The component client is installed successfully
   ```

NOTE

- If the clients of all or some services use the **/opt/client** directory, other directories must be used when you install other service clients.
- You must delete the client installation directory when uninstalling a client.
- To ensure that an installed client can only be used by the installation user (for example, **user_client**), add parameter **-o** during the installation. That is, run the **./install.sh /opt/client -o** command to install the client.
- If an HBase client is installed, it is recommended that the client installation directory contain only uppercase and lowercase letters, digits, and characters (_-?.@+=) due to the limitation of the Ruby syntax used by HBase.

Using a Client

1. On the node where the client is installed, run the **sudo su - omm** command to switch the user. Run the following command to go to the client directory:

   cd /opt/client

2. Run the following command to configure environment variables:

   source bigdata_env

3. If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If Kerberos authentication is disabled for the current cluster, skip this step.

   kinit *MRS cluster user*

   Example: **kinit admin**

   NOTE

   > User **admin** is created by default for MRS clusters with Kerberos authentication enabled and is used for administrators to maintain the clusters.

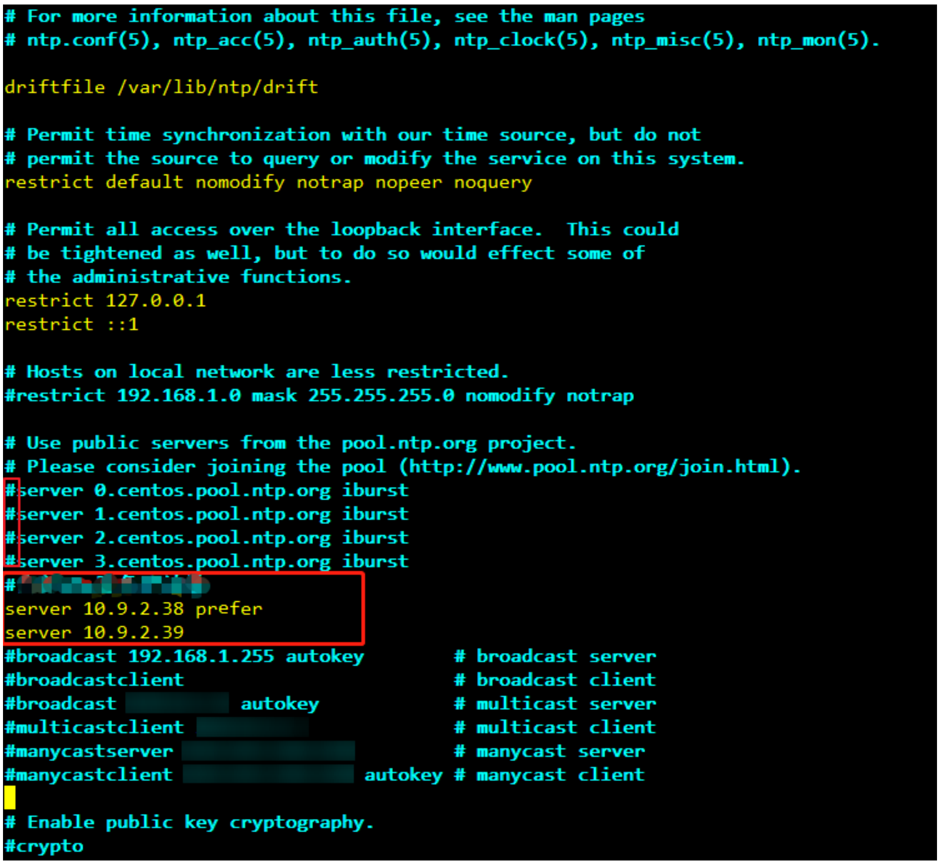
4. Run the client command of a component directly.

   For example, run the **hdfs dfs -ls /** command to view files in the HDFS root directory.

Installing a Client on a Node Outside a Cluster

1. Create an ECS that meets the requirements in **Prerequisites**.
2. Perform NTP time synchronization to synchronize the time of nodes outside the cluster with that of the MRS cluster.
   a. Run the **vi /etc/ntp.conf** command to edit the NTP client configuration file, add the IP addresses of the master node in the MRS cluster, and comment out the IP address of other servers.
   ```
   server master1_ip prefer
   server master2_ip
   ```

Figure 5-1 Adding the master node IP addresses



```
# For more information about this file, see the man pages
# ntp.conf(5), ntp_acc(5), ntp_auth(5), ntp_clock(5), ntp_misc(5), ntp_mon(5).

driftfile /var/lib/ntp/drift

# Permit time synchronization with our time source, but do not
# permit the source to query or modify the service on this system.
restrict default nomodify notrap nopeer noquery

# Permit all access over the loopback interface.  This could
# be tightened as well, but to do so would effect some of
# the administrative functions.
restrict 127.0.0.1
restrict ::1

# Hosts on local network are less restricted.
#restrict 192.168.1.0 mask 255.255.255.0 nomodify notrap

# Use public servers from the pool.ntp.org project.
# Please consider joining the pool (http://www.pool.ntp.org/join.html).
#server 0.centos.pool.ntp.org iburst
#server 1.centos.pool.ntp.org iburst
#server 2.centos.pool.ntp.org iburst
#server 3.centos.pool.ntp.org iburst
#[redacted]
server 10.9.2.38 prefer
server 10.9.2.39
#broadcast 192.168.1.255 autokey        # broadcast server
#broadcastclient                        # broadcast client
#broadcast [redacted]  autokey          # multicast server
#multicastclient [redacted]             # multicast client
#manycastserver [redacted]              # manycast server
#manycastclient [redacted]      autokey # manycast client

# Enable public key cryptography.
#crypto
```

b. Run the **service ntpd stop** command to stop the NTP service.

c. Run the following command to manually synchronize the time:

/usr/sbin/ntpdate *192.168.10.8*

NOTE

192.168.10.8 indicates the IP address of the active Master node.

d. Run the **service ntpd start** or **systemctl restart ntpd** command to start the NTP service.

e. Run the **ntpstat** command to check the time synchronization result.

3. Perform the following steps to download the cluster client software package from FusionInsight Manager, copy the package to the ECS node, and install the client:

a. Log in to FusionInsight Manager and download the cluster client to the specified directory on the active management node by referring to **Accessing FusionInsight Manager (MRS 3.x or Later)** and **Installing a Client on a Node Inside a Cluster**.

b. Log in to the active management node as user **root** and run the following command to copy the client installation package to the target node:

scp -p /tmp/FusionInsight-Client/FusionInsight_Cluster_1_Services_Client.tar *IP address of the node where the client is to be installed*:**/tmp**

c. Log in to the node on which the client is to be installed as the client user.

Run the following commands to install the client. If the user does not have operation permissions on the client software package and client installation directory, grant the permissions using the **root** user.

cd /tmp

tar -xvf FusionInsight_Cluster_1_Services_Client.tar

tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar

cd FusionInsight_Cluster_1_Services_ClientConfig

./install.sh /opt/client

d. Run the following commands to switch to the client directory and configure environment variables:

cd /opt/client

source bigdata_env

e. If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If Kerberos authentication is disabled for the current cluster, skip this step.

kinit *MRS cluster user*

Example: **kinit admin**

f. Run the client command of a component directly.

For example, run the **hdfs dfs -ls /** command to view files in the HDFS root directory.

5.1.2 Installing a Client (Versions Earlier Than 3.x)

Scenario

An MRS client is required. The MRS cluster client can be installed on the Master or Core node in the cluster or on a node outside the cluster.

After a cluster of versions earlier than MRS 3.x is created, a client is installed on the active Master node by default. You can directly use the client. The installation directory is **/opt/client**.

For details about how to install a client of MRS 3.x or later, see **Installing a Client (Version 3.x or Later)**.

NOTE

If a client has been installed on the node outside the MRS cluster and the client only needs to be updated, update the client using the user who installed the client, for example, user **root**.

Prerequisites

- An ECS has been prepared. For details about the OS and its version of the ECS, see **Table 5-2**.

Table 5-2 Reference list

OS	Supported Version
EulerOS	● Available: EulerOS 2.2 ● Available: EulerOS 2.3 ● Available: EulerOS 2.5

In addition, sufficient disk space is allocated for the ECS, for example, 40 GB.

- The ECS and the MRS cluster are in the same VPC.
- The security group of the ECS is the same as that of the Master node of the MRS cluster.

  If this requirement is not met, modify the ECS security group or configure the inbound and outbound rules of the ECS security group to allow the ECS security group to be accessed by all security groups of MRS cluster nodes.
- To enable users to log in to a Linux ECS using a password (SSH), see ***Instances*** > ***Logging In to a Linux ECS*** > ***Login Using an SSH Password*** *in the Elastic Cloud Server User Guide*.

Installing a Client on the Core Node

1. Log in to MRS Manager and choose **Services** > **Download Client** to download the client installation package to the active management node.

   NOTE

   If only the client configuration file needs to be updated, see method 2 in **Updating a Client (Versions Earlier Than 3.x)**.
2. Use the IP address to search for the active management node, and log in to the active management node using VNC.
3. Log in to the active management node, and run the following command to switch the user:

   sudo su - omm
4. On the MRS management console, view the IP address on the **Nodes** tab page of the specified cluster.

   Record the IP address of the Core node where the client is to be used.
5. On the active management node, run the following command to copy the client installation package to the Core node:

   scp -p /tmp/MRS-client/MRS_Services_Client.tar *IP address of the Core node*:**/opt/client**
6. Log in to the Core node as user **root**.

   For details, see **Login Using an SSH Key**.
7. Run the following commands to install the client:

   cd /opt/client

   tar -xvf MRS_Services_Client.tar

   tar -xvf MRS_Services_ClientConfig.tar

   cd /opt/client/MRS_Services_ClientConfig

./install.sh *Client installation directory*

For example, run the following command:

./install.sh /opt/client

8. For details about how to use the client, see **Using an MRS Client**.

Using an MRS Client

1. On the node where the client is installed, run the **sudo su - omm** command to switch the user. Run the following command to go to the client directory:

   cd /opt/client

2. Run the following command to configure environment variables:

   source bigdata_env

3. If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If Kerberos authentication is disabled for the current cluster, skip this step.

   kinit *MRS cluster user*

   Example: **kinit admin**

   NOTE

   User **admin** is created by default for MRS clusters with Kerberos authentication enabled and is used for administrators to maintain the clusters.

4. Run the client command of a component directly.

   For example, run the **hdfs dfs -ls /** command to view files in the HDFS root directory.

Installing a Client on a Node Outside the Cluster

Step 1 Create an ECS that meets the requirements in the prerequisites.

Step 2 Log in to MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**. Then, choose **Services**.

Step 3 Click **Download Client**.

Step 4 In **Client Type**, select **All client files**.

Step 5 In **Download To**, select **Remote host**.

Step 6 Set **Host IP Address** to the IP address of the ECS, **Host Port** to **22**, and **Save Path** to **/tmp**.

- If the default port **22** for logging in to an ECS using SSH has been changed, set **Host Port** to the new port.
- **Save Path** contains a maximum of 256 characters.

Step 7 Set **Login User** to **root**.

If other users are used, ensure that the users have read, write, and execute permission on the save path.

Step 8 Click **OK** to generate a client file.

If the following information is displayed, the client package is saved. Click **Close**. Obtain the client file from the save path on the remote host that is set when the client is downloaded.

```
Client files downloaded to the remote host successfully.
```

If the following information is displayed, check the username, password, and security group configurations of the remote host. Ensure that the username and password are correct and an inbound rule of the SSH (22) port has been added to the security group of the remote host. And then, go to **Step 2** to download the client again.

```
Failed to connect to the server. Please check the network connection or parameter settings.
```

NOTE

Generating a client will occupy a large number of disk I/Os. You are advised not to download a client when the cluster is being installed, started, and patched, or in other unstable states.

Step 9 Log in to the ECS using VNC. For details, see **Instance** > **Logging In to a Linux** > **Logging In to a Linux** in the *Elastic Cloud Server User Guide*

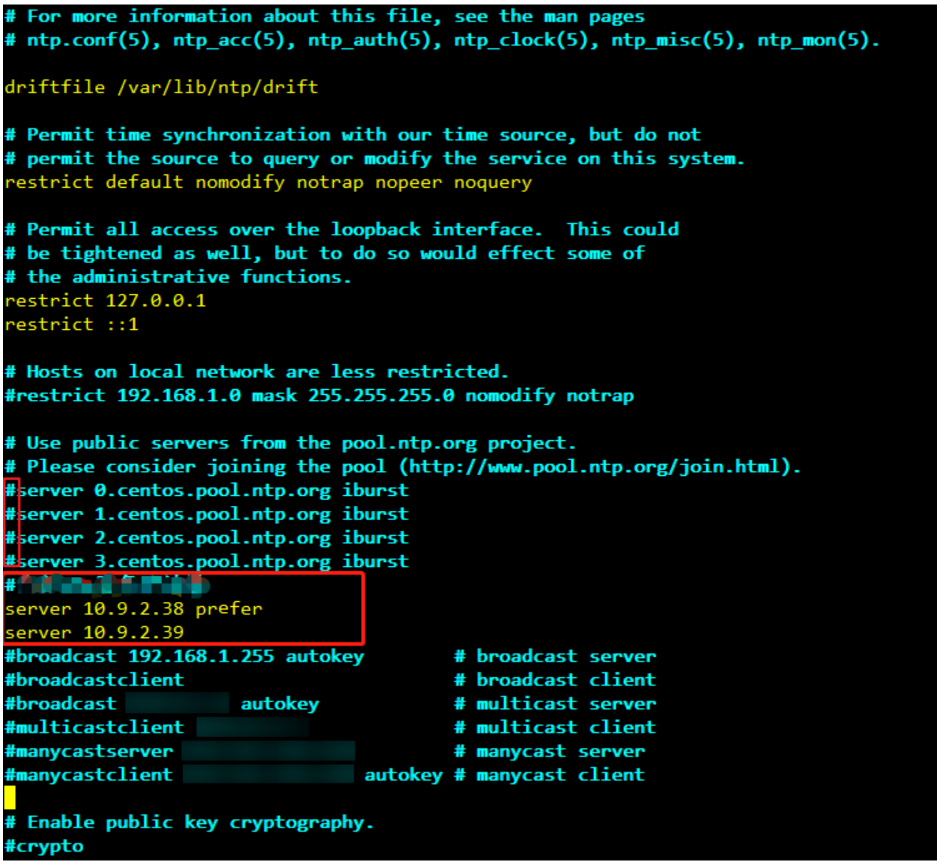
Log in to the ECS. For details, see **Login Using an SSH Key**. Set the ECS password and log in to the ECS in VNC mode.

Step 10 Perform NTP time synchronization to synchronize the time of nodes outside the cluster with the time of the MRS cluster.

1. Check whether the NTP service is installed. If it is not installed, run the **yum install ntp -y** command to install it.
2. Run the **vim /etc/ntp.conf** command to edit the NTP client configuration file, add the IP address of the Master node in the MRS cluster, and comment out the IP addresses of other servers.

```
server master1_ip prefer
server master2_ip
```

Figure 5-2 Adding the Master node IP addresses



```
# For more information about this file, see the man pages
# ntp.conf(5), ntp_acc(5), ntp_auth(5), ntp_clock(5), ntp_misc(5), ntp_mon(5).

driftfile /var/lib/ntp/drift

# Permit time synchronization with our time source, but do not
# permit the source to query or modify the service on this system.
restrict default nomodify notrap nopeer noquery

# Permit all access over the loopback interface.  This could
# be tightened as well, but to do so would effect some of
# the administrative functions.
restrict 127.0.0.1
restrict ::1

# Hosts on local network are less restricted.
#restrict 192.168.1.0 mask 255.255.255.0 nomodify notrap

# Use public servers from the pool.ntp.org project.
# Please consider joining the pool (http://www.pool.ntp.org/join.html).
#server 0.centos.pool.ntp.org iburst
#server 1.centos.pool.ntp.org iburst
#server 2.centos.pool.ntp.org iburst
#server 3.centos.pool.ntp.org iburst
#[redacted]
server 10.9.2.38 prefer
server 10.9.2.39
#broadcast 192.168.1.255 autokey        # broadcast server
#broadcastclient                        # broadcast client
#broadcast [redacted] autokey           # multicast server
#multicastclient [redacted]             # multicast client
#manycastserver [redacted]              # manycast server
#manycastclient [redacted] autokey # manycast client

# Enable public key cryptography.
#crypto
```

3. Run the **service ntpd stop** command to stop the NTP service.
4. Run the following command to manually synchronize the time:

   /usr/sbin/ntpdate *192.168.10.8*

   NOTE

   192.168.10.8 indicates the IP address of the active Master node.

5. Run the **service ntpd start** or **systemctl restart ntpd** command to start the NTP service.
6. Run the **ntpstat** command to check the time synchronization result:

Step 11 On the ECS, switch to user **root** and copy the installation package in **Save Path** in Step 6 to the **/opt** directory. For example, if **Save Path** is set to **/tmp**, run the following commands:

sudo su - root

cp /tmp/MRS_Services_Client.tar /opt

Step 12 Run the following command in the **/opt** directory to decompress the package and obtain the verification file and the configuration package of the client:

tar -xvf MRS_Services_Client.tar

Step 13 Run the following command to verify the configuration file package of the client:

sha256sum -c MRS_Services_ClientConfig.tar.sha256

The command output is as follows:

```
MRS_Services_ClientConfig.tar: OK
```

Step 14 Run the following command to decompress **MRS_Services_ClientConfig.tar**:

tar -xvf MRS_Services_ClientConfig.tar

Step 15 Run the following command to install the client to a new directory, for example, **/opt/Bigdata/client**. A directory is automatically generated during the client installation.

sh /opt/MRS_Services_ClientConfig/install.sh /opt/Bigdata/client

If the following information is displayed, the client has been successfully installed:

```
Components client installation is complete.
```

Step 16 Check whether the IP address of the ECS node is connected to the IP address of the cluster Master node.

For example, run the following command: **ping** *Master node IP address*.

- If yes, go to **Step 17**.
- If no, check whether the VPC and security group are correct and whether the ECS and the MRS cluster are in the same VPC and security group, and go to **Step 17**.

Step 17 Run the following command to configure environment variables:

source /opt/Bigdata/client/bigdata_env

Step 18 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If Kerberos authentication is disabled for the current cluster, skip this step.

kinit *MRS cluster user*

Example: **kinit admin**

Step 19 Run the client command of a component.

For example, run the following command to query the HDFS directory:

hdfs dfs -ls /

----End

5.2 Updating a Client

5.2.1 Updating a Client (Version 3.x or Later)

A cluster provides a client for you to connect to a server, view task results, or manage data. If you modify service configuration parameters on Manager and restart the service, you need to download and install the client again or use the configuration file to update the client.

Updating the Client Configuration

Method 1:

Step 1 Log in to FusionInsight Manager. For details, see **Accessing FusionInsight Manager (MRS 3.x or Later)**. Click the name of the cluster to be operated in the **Cluster** drop-down list.

Step 2 Choose **More** > **Download Client** > **Configuration Files Only**.

The generated compressed file contains the configuration files of all services.

Step 3 Determine whether to generate a configuration file on the cluster node.

- If yes, select **Save to Path**, and click **OK** to generate the client file. By default, the client file is generated in **/tmp/FusionInsight-Client** on the active management node. You can also store the client file in other directories, and user **omm** has the read, write, and execute permissions on the directories. Then go to **Step 4**.
- If no, click **OK**, specify a local save path, and download the complete client. Wait until the download is complete and go to **Step 4**.

Step 4 Use WinSCP to save the compressed file to the client installation directory, for example, **/opt/hadoopclient**, as the client installation user.

Step 5 Decompress the software package.

Run the following commands to go to the directory where the client is installed, and decompress the file to a local directory. For example, the downloaded client file is **FusionInsight_Cluster_1_Services_Client.tar**.

cd /opt/hadoopclient

tar -xvf FusionInsight_Cluster_1_Services_Client.tar

Step 6 Verify the software package.

Run the following command to verify the decompressed file and check whether the command output is consistent with the information in the **sha256** file.

sha256sum -c FusionInsight_Cluster_1_Services_ClientConfig_ConfigFiles.tar.sha256

```
FusionInsight_Cluster_1_Services_ClientConfig_ConfigFiles.tar: OK
```

Step 7 Decompress the package to obtain the configuration file.

tar -xvf FusionInsight_Cluster_1_Services_ClientConfig_ConfigFiles.tar

Step 8 Run the following command in the client installation directory to update the client using the configuration file:

sh refreshConfig.sh *Client installation directory Directory where the configuration file is located*

For example, run the following command:

sh refreshConfig.sh /opt/hadoopclient /opt/hadoopclient/FusionInsight_Cluster_1_Services_ClientConfig_ConfigFiles

If the following information is displayed, the configurations have been updated successfully.

```
Succeed to refresh components client config.
```

----End

Method 2:

Step 1 Log in to the client installation node as user **root**.

Step 2 Go to the client installation directory, for example, **/opt/hadoopclient** and run the following commands to update the configuration file:

cd /opt/hadoopclient

sh autoRefreshConfig.sh

Step 3 Enter the username and password of the FusionInsight Manager administrator and the floating IP address of FusionInsight Manager.

Step 4 Enter the names of the components whose configuration needs to be updated. Use commas (,) to separate the component names. Press **Enter** to update the configurations of all components if necessary.

If the following information is displayed, the configurations have been updated successfully.

```
Succeed to refresh components client config.
```

----End

5.2.2 Updating a Client (Versions Earlier Than 3.x)

NOTE

This section applies to clusters of versions earlier than MRS 3.x. For MRS 3.x or later, see **Updating a Client (Version 3.x or Later)**.

Updating a Client Configuration File

Scenario

An MRS cluster provides a client for you to connect to a server, view task results, or manage data. Before using an MRS client, you need to download and update the client configuration file if service configuration parameters are modified and a service is restarted or the service is merely restarted on MRS Manager.

During cluster creation, the original client is stored in the **/opt/client** directory on all nodes in the cluster by default. After the cluster is created, only the client of a Master node can be directly used. To use the client of a Core node, you need to update the client configuration file first.

Procedure

Method 1:

Step 1 Log in to MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**. Then, choose **Services**.

Step 2 Click **Download Client**.

Set **Client Type** to **Only configuration files**, **Download To** to **Server**, and click **OK** to generate the client configuration file. The generated file is saved in the **/tmp/MRS-client** directory on the active management node by default. You can customize the file path.

Step 3 Query and log in to the active Master node.

Step 4 If you use the client in the cluster, run the following command to switch to user **omm**. If you use the client outside the cluster, switch to user **root**.

sudo su - omm

Step 5 Run the following command to switch to the client directory, for example, **/opt/Bigdata/client**:

cd /opt/Bigdata/client

Step 6 Run the following command to update client configurations:

sh refreshConfig.sh *Client installation directory Full path of the client configuration file package*

For example, run the following command:

sh refreshConfig.sh /opt/Bigdata/client /tmp/MRS-client/MRS_Services_Client.tar

If the following information is displayed, the configurations have been updated successfully.

```
ReFresh components client config is complete.
Succeed to refresh components client config.
```

----End

Method 2:

Step 1 After the cluster is installed, run the following command to switch to user **omm**. If you use the client outside the cluster, switch to user **root**.

sudo su - omm

Step 2 Run the following command to switch to the client directory, for example, **/opt/Bigdata/client**:

cd /opt/Bigdata/client

Step 3 Run the following command and enter the name of an MRS Manager user with the download permission and its password (for example, the username is **admin** and the password is the one set during cluster creation) as prompted to update client configurations.

sh autoRefreshConfig.sh

Step 4 After the command is executed, the following information is displayed, where *XXX* indicates the name of the component installed in the cluster. To update client configurations of all components, press **Enter**. To update client configurations of some components, enter the component names and separate them with commas (,).

```
Components "xxx" have been installed in the cluster. Please input the comma-separated names of the components for which you want to update client configurations. If you press Enter without inputting any component name, the client configurations of all components will be updated:
```

If the following information is displayed, the configurations have been updated successfully.

```
Succeed to refresh components client config.
```

If the following information is displayed, the username or password is incorrect.

```
login manager failed,Incorrect username or password.
```

NOTE

- This script automatically connects to the cluster and invokes the **refreshConfig.sh** script to download and update the client configuration file.
- By default, the client uses the floating IP address specified by **wsom=xxx** in the **Version** file in the installation directory to update the client configurations. To update the configuration file of another cluster, modify the value of **wsom=xxx** in the **Version** file to the floating IP address of the corresponding cluster before performing this step.

----End

Fully Updating the Original Client of the Active Master Node

Scenario

During cluster creation, the original client is stored in the **/opt/client** directory on all nodes in the cluster by default. The following uses **/opt/Bigdata/client** as an example.

- For a normal MRS cluster, you will use the pre-installed client on a Master node to submit a job on the management console page.
- You can also use the pre-installed client on the Master node to connect to a server, view task results, and manage data.

After installing the patch on the cluster, you need to update the client on the Master node to ensure that the functions of the built-in client are available.

Procedure

Step 1 Log in to MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**. Then, choose **Services**.

Step 2 Click **Download Client**.

Set **Client Type** to **All client files**, **Download To** to **Server**, and click **OK** to generate the client configuration file. The generated file is saved in the **/tmp/MRS-client** directory on the active management node by default. You can customize the file path.

Step 3 Query and log in to the active Master node.

Step 4 On the ECS, switch to user **root** and copy the installation package to the **/opt** directory.

sudo su - root

cp /tmp/MRS-client/MRS_Services_Client.tar /opt

Step 5 Run the following command in the **/opt** directory to decompress the package and obtain the verification file and the configuration package of the client:

tar -xvf MRS_Services_Client.tar

Step 6 Run the following command to verify the configuration file package of the client:

sha256sum -c MRS_Services_ClientConfig.tar.sha256

The command output is as follows:

```
MRS_Services_ClientConfig.tar: OK
```

Step 7 Run the following command to decompress **MRS_Services_ClientConfig.tar**:

tar -xvf MRS_Services_ClientConfig.tar

Step 8 Run the following command to move the original client to the **/opt/Bigdata/client_bak** directory:

mv /opt/Bigdata/client /opt/Bigdata/client_bak

Step 9 Run the following command to install the client in a new directory. The client path must be **/opt/Bigdata/client**.

sh /opt/MRS_Services_ClientConfig/install.sh /opt/Bigdata/client

If the following information is displayed, the client has been successfully installed:

```
Components client installation is complete.
```

Step 10 Run the following command to modify the user and user group of the **/opt/Bigdata/client** directory:

chown omm:wheel /opt/Bigdata/client -R

Step 11 Run the following command to configure environment variables:

source /opt/Bigdata/client/bigdata_env

Step 12 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If Kerberos authentication is disabled for the current cluster, skip this step.

kinit *MRS cluster user*

Example: **kinit admin**

Step 13 Run the client command of a component.

For example, run the following command to query the HDFS directory:

hdfs dfs -ls /

----End

Fully Updating the Original Client of the Standby Master Node

Step 1 Repeat **Step 1** to **Step 3** to log in to the standby Master node, and run the following command to switch to user **omm**:

sudo su - omm

Step 2 Run the following command on the standby master node to copy the downloaded client package from the active master node:

scp omm@*master1 nodeIP address***:/tmp/MRS-client/MRS_Services_Client.tar /tmp/MRS-client/**

NOTE

- In this command, **master1** node is the active master node.
- **/tmp/MRS-client/** is an example target directory of the standby master node.

Step 3 Repeat **Step 4** to **Step 13** to update the client of the standby Master node.

----End

5.3 Using the Client of Each Component

5.3.1 Using a ClickHouse Client

ClickHouse is a column-based database oriented to online analysis and processing. It supports SQL query and provides good query performance. The aggregation analysis and query performance based on large and wide tables is excellent, which is one order of magnitude faster than other analytical databases.

Prerequisites

You have installed the client, for example, in the **/opt/hadoopclient** directory. The client directory in the following operations is only an example. Change it to the actual installation directory. Before using the client, download and update the client configuration file, and ensure that the active management node of Manager is available.

Procedure

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client installation directory:

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the current user. The current user must have the permission to create ClickHouse tables. For details about how to bind a role to the user, see . If Kerberos authentication is disabled for the current cluster, skip this step.

1. Run the following command if it is an MRS 3.1.0 cluster:

   export CLICKHOUSE_SECURITY_ENABLED=true
2. **kinit** *Component service user*

   Example: **kinit clickhouseuser**

Step 5 Run the client command of the ClickHouse component.

Run the **clickhouse -h** command to view the command help of ClickHouse.

The command output is as follows:

```
Use one of the following commands:
clickhouse local [args]
clickhouse client [args]
clickhouse benchmark [args]
clickhouse server [args]
clickhouse performance-test [args]
clickhouse extract-from-config [args]
clickhouse compressor [args]
clickhouse format [args]
clickhouse copier [args]
clickhouse obfuscator [args]
...
```

For details about how to use the command, see **https://clickhouse.tech/docs/en/operations/**.

Run the **clickhouse client** command to connect to the ClickHouse serverif MRS 3.1.0 or later.

- Command for using SSL to log in to a ClickHouse cluster with Kerberos authentication disabled

  clickhouse client --host *IP address of the ClickHouse instance***--user** *Username* **--password --port** 9440 **--secure**

  Enter the user password.

- Using SSL for login when Kerberos authentication is enabled for the current cluster:

  You must create a user on Manager because there is no default user. For details, see .

  After the user authentication is successful, you do not need to carry the **--user** and **--password** parameters when logging in to the client as the authenticated user.

  clickhouse client --host *IP address of the ClickHouse instance* **--port** 9440 **--secure**

  The following table describes the parameters of the **clickhouse client** command.

Table 5-3 Parameters of the **clickhouse client** command

Parameter	Description
--host	Host name of the server. The default value is **localhost**. You can use the host name or IP address of the node where the ClickHouse instance is located. **NOTE** You can log in to FusionInsight Manager and choose **Cluster** > **Services** > **ClickHouse** > **Instance** to obtain the service IP address of the ClickHouseServer instance.

Parameter	Description
--port	Port for connection. ● If the SSL security connection is used, the default port number is **9440**, the parameter **--secure** must be carried. For details about the port number, search for the **tcp_port_secure** parameter in the ClickHouseServer instance configuration. ● If non-SSL security connection is used, the default port number is **9000**, the parameter **--secure** does not need to be carried. For details about the port number, search for the **tcp_port** parameter in the ClickHouseServer instance configuration.
--user	Username. You can create the user on Manager and bind a role to the user. For details, see . ● If Kerberos authentication is enabled for the current cluster and the user authentication is successful, you do not need to carry the **--user** and **--password** parameters when logging in to the client as the authenticated user. You must create a user with this name on Manager because there is no default user in the Kerberos cluster scenario. ● If Kerberos authentication is not enabled for the current cluster, you can specify a user and its password created on Manager when logging in to the client. If the user and password parameters are not carried, user **default** is used for login by default. The user in normal mode (Kerberos authentication disabled) is the default user, or you can create an administrator using the open source capability provided by the ClickHouse community. You cannot use the users created on FusionInsight Manager.
--password	Password. The default password is an empty string. This parameter is used together with the **--user** parameter. You can set a password when creating a user on Manager.
--query	Query to process when using non-interactive mode.
--database	Current default database. The default value is **default**, which is the default configuration on the server.
--multiline	If this parameter is specified, multiline queries are allowed. (**Enter** only indicates line feed and does not indicate that the query statement is complete.)
--multiquery	If this parameter is specified, multiple queries separated with semicolons (;) can be processed. This parameter is valid only in non-interactive mode.
--format	Specified default format used to output the result.
--vertical	If this parameter is specified, the result is output in vertical format by default. In this format, each value is printed on a separate line, which helps to display a wide table.

Parameter	Description
--time	If this parameter is specified, the query execution time is printed to **stderr** in non-interactive mode.
--stacktrace	If this parameter is specified, stack trace information will be printed when an exception occurs.
--config-file	Name of the configuration file.
--secure	If this parameter is specified, the server will be connected in SSL mode.
--history_file	Path of files that record command history.
--param_<name>	Query with parameters. Pass values from the client to the server. For details, see **https://clickhouse.tech/docs/en/interfaces/cli/#cli-queries-with-parameters**.

----End

5.3.2 Using a Flink Client

This section describes how to use Flink to run wordcount jobs.

Prerequisites

- Flink has been installed in an MRS cluster.
- The cluster runs properly and the client has been correctly installed, for example, in the **/opt/hadoopclient** directory. The client directory in the following operations is only an example. Change it to the actual installation directory.

Using the Flink Client (Versions Earlier Than MRS 3.x)

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client installation directory:

cd /opt/hadoopclient

Step 3 Run the following command to initialize environment variables:

source /opt/hadoopclient/bigdata_env

Step 4 If Kerberos authentication is enabled for the cluster, perform the following steps. If not, skip this whole step.

1. Prepare a user for submitting Flink jobs..
2. Log in to Manager and download the authentication credential.

   Log in to Manager of the cluster. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**. Choose **System Settings** > **User Management**. In the **Operation** column of the row that contains the added user, choose **More** > **Download Authentication Credential**.

3. Decompress the downloaded authentication credential package and copy the **user.keytab** file to the client node, for example, to the **/opt/hadoopclient/Flink/flink/conf** directory on the client node. If the client is installed on a node outside the cluster, copy the **krb5.conf** file to the **/etc/** directory on this node.
4. Configure security authentication by adding the **keytab** path and username in the **/opt/hadoopclient/Flink/flink/conf/flink-conf.yaml** configuration file.

   security.kerberos.login.keytab: *<user.keytab file path>*

   security.kerberos.login.principal: *<Username>*

   Example:

   security.kerberos.login.keytab: /opt/hadoopclient/Flink/flink/conf/user.keytab

   security.kerberos.login.principal: test
5. In the **bin** directory of the Flink client, run the following command to perform security hardening. For details, see . Set **password** in the following command to a password for submitting jobs:

   sh generate_keystore.sh <*password*>

   The script automatically replaces the SSL value in the **/opt/hadoopclient/Flink/flink/conf/flink-conf.yaml** file. For an MRS 2.*x* or earlier security cluster, external SSL is disabled by default. To enable external SSL, configure the parameter and run the script again. For details, see .

   NOTE

   - You do not need to manually generate the **generate_keystore.sh** script.
   - After authentication and encryption, the generated **flink.keystore**, **flink.truststore**, and **security.cookie** items are automatically filled in the corresponding configuration items in **flink-conf.yaml**.
6. Configure paths for the client to access the **flink.keystore** and **flink.truststore** files.
   - Absolute path: After the script is executed, the file path of **flink.keystore** and **flink.truststore** is automatically set to the absolute path **/opt/hadoopclient/Flink/flink/conf/** in the **flink-conf.yaml** file. In this case, you need to move the **flink.keystore** and **flink.truststore** files from the **conf** directory to this absolute path on the Flink client and Yarn nodes.
   - Relative path: Perform the following steps to set the file path of **flink.keystore** and **flink.truststore** to the relative path and ensure that the directory where the Flink client command is executed can directly access the relative paths.
     i. Create a directory, for example, **ssl**, in **/opt/hadoopclient/Flink/flink/conf/**.

        cd /opt/hadoopclient/Flink/flink/conf/

        mkdir ssl

     ii. Move the **flink.keystore** and **flink.truststore** files to the **/opt/hadoopclient/Flink/flink/conf/ssl/** directory.

        mv flink.keystore ssl/

        mv flink.truststore ssl/

     iii. Change the values of the following parameters to relative paths in the **flink-conf.yaml** file:

```
security.ssl.internal.keystore: ssl/flink.keystore
security.ssl.internal.truststore: ssl/flink.truststore
```

Step 5 Run a wordcount job.

NOTICE

To submit or run jobs on Flink, the user must have the following permissions:

- If Ranger authentication is enabled, the current user must belong to the **hadoop** group or the user has been granted the **/flink** read and write permissions in Ranger.
- If Ranger authentication is disabled, the current user must belong to the **hadoop** group.

- Normal cluster (Kerberos authentication disabled)
  - Run the following commands to start a session and submit a job in the session:

    yarn-session.sh -nm "*session-name***"**

    flink run /opt/hadoopclient/Flink/flink/examples/streaming/WordCount.jar
  - Run the following command to submit a single job on Yarn:

    flink run -m yarn-cluster /opt/hadoopclient/Flink/flink/examples/streaming/WordCount.jar
- Security cluster (Kerberos authentication enabled)
  - If the **flink.keystore** and **flink.truststore** file are stored in the absolute path:
    - Run the following commands to start a session and submit a job in the session:

      yarn-session.sh -nm "*session-name***"**

      flink run /opt/hadoopclient/Flink/flink/examples/streaming/WordCount.jar
    - Run the following command to submit a single job on Yarn:

      flink run -m yarn-cluster /opt/hadoopclient/Flink/flink/examples/streaming/WordCount.jar
  - If the **flink.keystore** and **flink.truststore** files are stored in the relative path:
    - In the same directory of SSL, run the following commands to start a session and submit jobs in the session. The SSL directory is a relative path. For example, if the SSL directory is **opt/hadoopclient/Flink/flink/conf/**, then run the following commands in this directory:

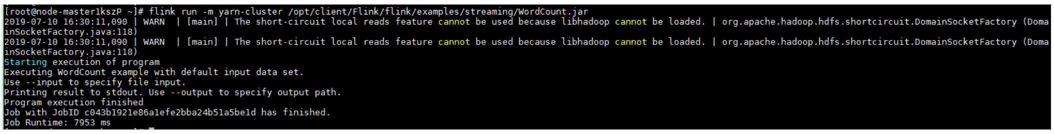
      yarn-session.sh -t ssl/ -nm "*session-name***"**

      flink run /opt/hadoopclient/Flink/flink/examples/streaming/WordCount.jar
    - Run the following command to submit a single job on Yarn:

flink run -m yarn-cluster -yt ssl/ /opt/hadoopclient/Flink/flink/examples/streaming/WordCount.jar

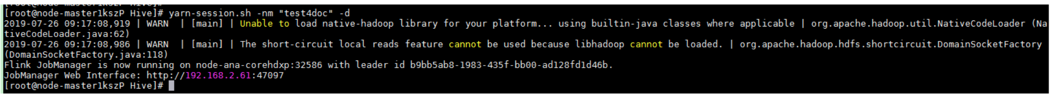
Step 6 After the job has been successfully submitted, the following information is displayed on the client:

Figure 5-3 Job submitted successfully on Yarn



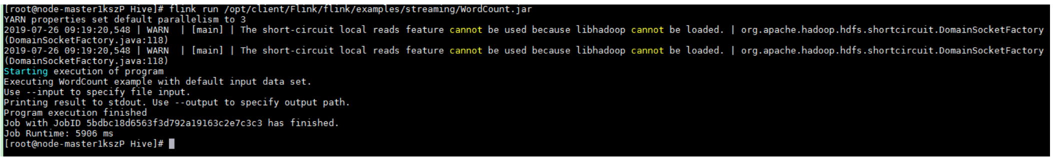
```
[root@node-master1kszP ~]# flink run -m yarn-cluster /opt/client/Flink/flink/examples/streaming/WordCount.jar
2019-07-10 16:30:11,090 | WARN  | [main] | The short-circuit local reads feature cannot be used because libhadoop cannot be loaded. | org.apache.hadoop.hdfs.shortcircuit.DomainSocketFactory (Doma
inSocketFactory.java:118)
2019-07-10 16:30:11,090 | WARN  | [main] | The short-circuit local reads feature cannot be used because libhadoop cannot be loaded. | org.apache.hadoop.hdfs.shortcircuit.DomainSocketFactory (Doma
inSocketFactory.java:118)
Starting execution of program
Executing WordCount example with default input data set.
Use --input to specify file input.
Printing result to stdout. Use --output to specify output path.
Program execution finished
Job with JobID c043b1921e86a1efe2bba24b51a5be1d has finished.
Job Runtime: 7953 ms
```

Figure 5-4 Session started successfully



```
[root@node-master1kszP Hive]# yarn-session.sh -nm "test4doc" -d
2019-07-26 09:17:08,919 | WARN  | [main] | Unable to load native-hadoop library for your platform... using builtin-java classes where applicable | org.apache.hadoop.util.NativeCodeLoader (Na
tiveCodeLoader.java:62)
2019-07-26 09:17:08,986 | WARN  | [main] | The short-circuit local reads feature cannot be used because libhadoop cannot be loaded. | org.apache.hadoop.hdfs.shortcircuit.DomainSocketFactory
(DomainSocketFactory.java:118)
Flink JobManager is now running on node-ana-corehdxp:32586 with leader id b9bb5ab8-1983-435f-bb00-ad128fd1d46b.
JobManager Web Interface: http://192.168.2.61:47097
[root@node-master1kszP Hive]#
```

Figure 5-5 Job submitted successfully in the session

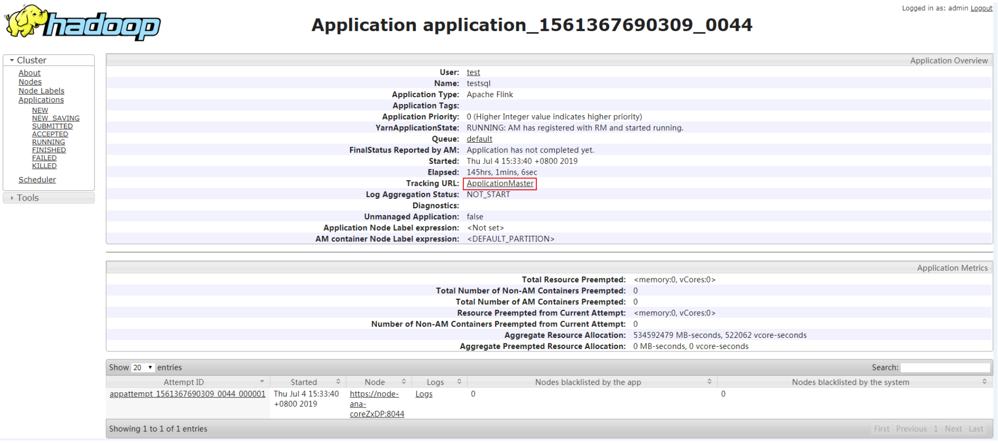


```
[root@node-master1kszP Hive]# flink run /opt/client/Flink/flink/examples/streaming/WordCount.jar
YARN properties set default parallelism to 3
2019-07-26 09:19:20,548 | WARN  | [main] | The short-circuit local reads feature cannot be used because libhadoop cannot be loaded. | org.apache.hadoop.hdfs.shortcircuit.DomainSocketFactory
(DomainSocketFactory.java:118)
2019-07-26 09:19:20,548 | WARN  | [main] | The short-circuit local reads feature cannot be used because libhadoop cannot be loaded. | org.apache.hadoop.hdfs.shortcircuit.DomainSocketFactory
(DomainSocketFactory.java:118)
Starting execution of program
Executing WordCount example with default input data set.
Use --input to specify file input.
Printing result to stdout. Use --output to specify output path.
Program execution finished
Job with JobID 5bdbc18d6563f3d792a19163c2e7c3c3 has finished.
Job Runtime: 5906 ms
[root@node-master1kszP Hive]#
```

Step 7 Go to the native YARN service page, find the application of the job, and click the application name to go to the job details page. For details, see .

- If the job is not completed, click **Tracking URL** to go to the native Flink page and view the job running information.
- If the job submitted in a session has been completed, you can click **Tracking URL** to log in to the native Flink service page to view job information.

Figure 5-6 Application



Application application_1561367690309_0044

Application Overview

User: test
Name: testsql
Application Type: Apache Flink
Application Tags:
Application Priority: 0 (Higher Integer value indicates higher priority)
YarnApplicationState: RUNNING: AM has registered with RM and started running.
Queue: default
FinalStatus Reported by AM: Application has not completed yet.
Started: Thu Jul 4 15:33:40 +0800 2019
Elapsed: 145hrs, 1mins, 6sec
Tracking URL: ApplicationMaster
Log Aggregation Status: NOT_START
Diagnostics:
Unmanaged Application: false
Application Node Label expression: <Not set>
AM container Node Label expression: <DEFAULT_PARTITION>

Application Metrics

Total Resource Preempted: <memory:0, vCores:0>
Total Number of Non-AM Containers Preempted: 0
Total Number of AM Containers Preempted: 0
Resource Preempted from Current Attempt: <memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt: 0
Aggregate Resource Allocation: 534592479 MB-seconds, 522062 vcore-seconds
Aggregate Preempted Resource Allocation: 0 MB-seconds, 0 vcore-seconds

Show 20 entries Search:

Attempt ID	Started	Node	Logs	Nodes blacklisted by the app	Nodes blacklisted by the system
appattempt_1561367690309_0044_000001	Thu Jul 4 15:33:40 +0800 2019	https://node-ana-coreZxDP:8044	Logs	0	0

Showing 1 to 1 of 1 entries

----End

Using the Flink Client (MRS 3.x or Later)

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client installation directory:

cd /opt/hadoopclient

Step 3 Run the following command to initialize environment variables:

source /opt/hadoopclient/bigdata_env

Step 4 If Kerberos authentication is enabled for the cluster, perform the following steps. If not, skip this whole step.

1. Prepare a user for submitting Flink jobs.
2. Log in to Manager and download the authentication credential.

   Log in to Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**. Choose **System** > **Permission** > **Manage User**. On the displayed page, locate the row that contains the added user, click **More** in the **Operation** column, and select **Download authentication credential**.
3. Decompress the downloaded authentication credential package and copy the **user.keytab** file to the client node, for example, to the **/opt/hadoopclient/Flink/flink/conf** directory on the client node. If the client is installed on a node outside the cluster, copy the **krb5.conf** file to the **/etc/** directory on this node.
4. Add the service IP address of the node where the client is installed and IP address of the master node to the **jobmanager.web.access-control-allow-origin** and **jobmanager.web.allow-access-address** configuration items in the **/opt/hadoopclient/Flink/flink/conf/flink-conf.yaml** file. Use commas (,) to separate IP addresses.

```
jobmanager.web.access-control-allow-origin: xx.xx.xxx.xxx,xx.xx.xxx.xxx,xx.xx.xxx.xxx
jobmanager.web.allow-access-address: xx.xx.xxx.xxx,xx.xx.xxx.xxx,xx.xx.xxx.xxx
```

   NOTE

   To obtain the service IP address of the node where the client is installed, perform the following operations:

   - Node inside the cluster:

     In the navigation tree of the MRS management console, choose **Clusters > Active Clusters**, select a cluster, and click its name to switch to the cluster details page.

     On the **Nodes** tab page, view the IP address of the node where the client is installed.
   - Node outside the cluster: IP address of the ECS where the client is installed.
5. Configure security authentication by adding the **keytab** path and username in the **/opt/hadoopclient/Flink/flink/conf/flink-conf.yaml** configuration file.

   security.kerberos.login.keytab: *<user.keytab file path>*

   security.kerberos.login.principal: *<Username>*

   Example:

   security.kerberos.login.keytab: /opt/hadoopclient/Flink/flink/conf/user.keytab

   security.kerberos.login.principal: test
6. Generate the **generate_keystore.sh** script and place it in the **bin** directory of the Flink client. For details, see . In the **bin** directory of the Flink client, run the following command to perform security hardening. For details, see . Set **password** in the following command to a password for submitting jobs:

   sh generate_keystore.sh <*password*>

The script automatically replaces the SSL value in the **/opt/hadoopclient/Flink/flink/conf/flink-conf.yaml** file.

sh generate_keystore.sh <*password*>

NOTE

After authentication and encryption, the **flink.keystore** and **flink.truststore** files are generated in the **conf** directory on the Flink client and the following configuration items are set to the default values in the **flink-conf.yaml** file:

- Set **security.ssl.keystore** to the absolute path of the **flink.keystore** file.
- Set **security.ssl.truststore** to the absolute path of the **flink.truststore** file.
- Set **security.cookie** to a random password automatically generated by the **generate_keystore.sh** script.
- By default, **security.ssl.encrypt.enabled** is set to **false** in the **flink-conf.yaml** file by default. The **generate_keystore.sh** script sets **security.ssl.key-password**, **security.ssl.keystore-password**, and **security.ssl.truststore-password** to the password entered when the **generate_keystore.sh** script is called.

7. Configure paths for the client to access the **flink.keystore** and **flink.truststore** files.
   - Absolute path: After the script is executed, the file path of **flink.keystore** and **flink.truststore** is automatically set to the absolute path **/opt/hadoopclient/Flink/flink/conf/** in the **flink-conf.yaml** file. In this case, you need to move the **flink.keystore** and **flink.truststore** files from the **conf** directory to this absolute path on the Flink client and Yarn nodes.
   - Relative path: Perform the following steps to set the file path of **flink.keystore** and **flink.truststore** to the relative path and ensure that the directory where the Flink client command is executed can directly access the relative paths.
     i. Create a directory, for example, **ssl**, in **/opt/hadoopclient/Flink/flink/conf/**.

        cd /opt/hadoopclient/Flink/flink/conf/

        mkdir ssl

     ii. Move the **flink.keystore** and **flink.truststore** files to the **/opt/hadoopclient/Flink/flink/conf/ssl/** directory.

        mv flink.keystore ssl/

        mv flink.truststore ssl/

     iii. Change the values of the following parameters to relative paths in the **flink-conf.yaml** file:

```
security.ssl.keystore: ssl/flink.keystore
security.ssl.truststore: ssl/flink.truststore
```

Step 5 Run a wordcount job.

NOTICE

To submit or run jobs on Flink, the user must have the following permissions:

- If Ranger authentication is enabled, the current user must belong to the **hadoop** group or the user has been granted the **/flink** read and write permissions in Ranger.
- If Ranger authentication is disabled, the current user must belong to the **hadoop** group.

- Normal cluster (Kerberos authentication disabled)
  - Run the following commands to start a session and submit a job in the session:

    yarn-session.sh -nm "*session-name***"**

    flink run /opt/hadoopclient/Flink/flink/examples/streaming/WordCount.jar
  - Run the following command to submit a single job on Yarn:

    flink run -m yarn-cluster /opt/hadoopclient/Flink/flink/examples/streaming/WordCount.jar
- Security cluster (Kerberos authentication enabled)
  - If the **flink.keystore** and **flink.truststore** files are stored in the absolute path:
    - Run the following commands to start a session and submit a job in the session:

      yarn-session.sh -nm "*session-name***"**

      flink run /opt/hadoopclient/Flink/flink/examples/streaming/WordCount.jar
    - Run the following command to submit a single job on Yarn:

      flink run -m yarn-cluster /opt/hadoopclient/Flink/flink/examples/streaming/WordCount.jar
  - If the **flink.keystore** and **flink.truststore** file are stored in the relative path:
    - In the same directory of SSL, run the following commands to start a session and submit jobs in the session. The SSL directory is a relative path. For example, if the SSL directory is **opt/hadoopclient/Flink/flink/conf/**, then run the following commands in this directory:

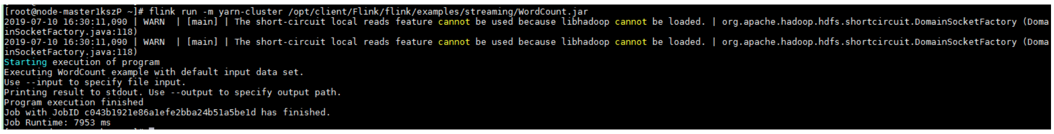
      yarn-session.sh -t ssl/ -nm "*session-name***"**

      flink run /opt/hadoopclient/Flink/flink/examples/streaming/WordCount.jar
    - Run the following command to submit a single job on Yarn:

      flink run -m yarn-cluster -yt ssl/ /opt/hadoopclient/Flink/flink/examples/streaming/WordCount.jar

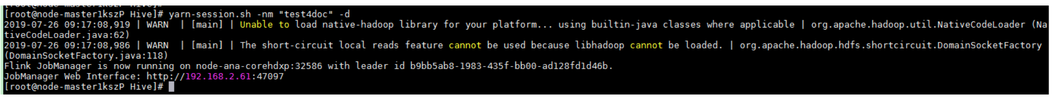
Step 6 After the job has been successfully submitted, the following information is displayed on the client:

Figure 5-7 Job submitted successfully on Yarn



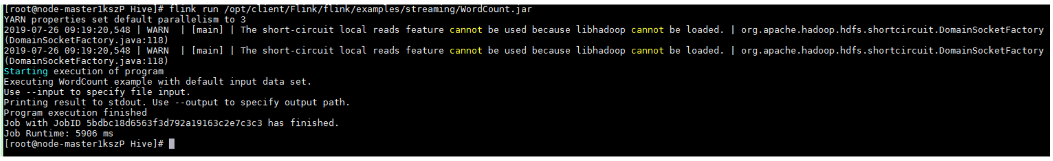
```
[root@node-master1kszP ~]# flink run -m yarn-cluster /opt/client/Flink/flink/examples/streaming/WordCount.jar
2019-07-10 16:30:11,090 | WARN  | [main] | The short-circuit local reads feature cannot be used because libhadoop cannot be loaded. | org.apache.hadoop.hdfs.shortcircuit.DomainSocketFactory (Doma
inSocketFactory.java:118)
2019-07-10 16:30:11,090 | WARN  | [main] | The short-circuit local reads feature cannot be used because libhadoop cannot be loaded. | org.apache.hadoop.hdfs.shortcircuit.DomainSocketFactory (Doma
inSocketFactory.java:118)
Starting execution of program
Executing WordCount example with default input data set.
Use --input to specify file input.
Printing result to stdout. Use --output to specify output path.
Program execution finished
Job with JobID c043b1921e86a1efe2bba24b51a5be1d has finished.
Job Runtime: 7953 ms
```

Figure 5-8 Session started successfully



```
[root@node-master1kszP Hive]# yarn-session.sh -nm "test4doc" -d
2019-07-26 09:17:08,919 | WARN  | [main] | Unable to load native-hadoop library for your platform... using builtin-java classes where applicable | org.apache.hadoop.util.NativeCodeLoader (Na
tiveCodeLoader.java:62)
2019-07-26 09:17:08,986 | WARN  | [main] | The short-circuit local reads feature cannot be used because libhadoop cannot be loaded. | org.apache.hadoop.hdfs.shortcircuit.DomainSocketFactory
(DomainSocketFactory.java:118)
Flink JobManager is now running on node-ana-corehdxp:32586 with leader id b9bb5ab8-1983-435f-bb00-ad128fd1d46b.
JobManager Web Interface: http://192.168.2.61:47097
[root@node-master1kszP Hive]#
```

Figure 5-9 Job submitted successfully in the session

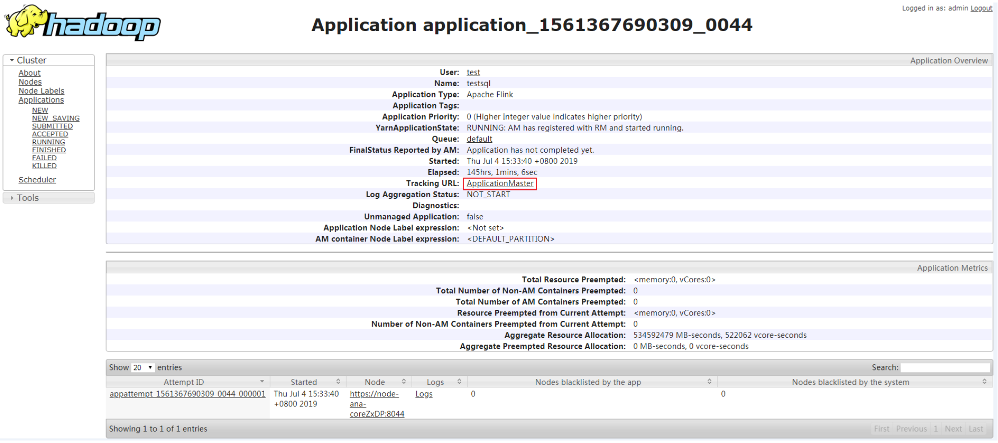


```
[root@node-master1kszP Hive]# flink run /opt/client/Flink/flink/examples/streaming/WordCount.jar
YARN properties set default parallelism to 3
2019-07-26 09:19:20,548 | WARN  | [main] | The short-circuit local reads feature cannot be used because libhadoop cannot be loaded. | org.apache.hadoop.hdfs.shortcircuit.DomainSocketFactory
(DomainSocketFactory.java:118)
2019-07-26 09:19:20,548 | WARN  | [main] | The short-circuit local reads feature cannot be used because libhadoop cannot be loaded. | org.apache.hadoop.hdfs.shortcircuit.DomainSocketFactory
(DomainSocketFactory.java:118)
Starting execution of program
Executing WordCount example with default input data set.
Use --input to specify file input.
Printing result to stdout. Use --output to specify output path.
Program execution finished
Job with JobID 5bdbc18d6563f3d792a19163c2e7c3c3 has finished.
Job Runtime: 5906 ms
[root@node-master1kszP Hive]#
```

Step 7 Go to the native YARN service page, find the application of the job, and click the application name to go to the job details page. For details, see .

- If the job is not completed, click **Tracking URL** to go to the native Flink page and view the job running information.
- If the job submitted in a session has been completed, you can click **Tracking URL** to log in to the native Flink service page to view job information.

Figure 5-10 Application



Application application_1561367690309_0044

Application Overview

User: test
Name: testsql
Application Type: Apache Flink
Application Tags:
Application Priority: 0 (Higher Integer value indicates higher priority)
YarnApplicationState: RUNNING: AM has registered with RM and started running.
Queue: default
FinalStatus Reported by AM: Application has not completed yet.
Started: Thu Jul 4 15:33:40 +0800 2019
Elapsed: 145hrs, 1mins, 6sec
Tracking URL: ApplicationMaster
Log Aggregation Status: NOT_START
Diagnostics:
Unmanaged Application: false
Application Node Label expression: <Not set>
AM container Node Label expression: <DEFAULT_PARTITION>

Application Metrics

Total Resource Preempted: <memory:0, vCores:0>
Total Number of Non-AM Containers Preempted: 0
Total Number of AM Containers Preempted: 0
Resource Preempted from Current Attempt: <memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt: 0
Aggregate Resource Allocation: 534592479 MB-seconds, 522062 vcore-seconds
Aggregate Preempted Resource Allocation: 0 MB-seconds, 0 vcore-seconds

Show 20 entries Search:

Attempt ID	Started	Node	Logs	Nodes blacklisted by the app	Nodes blacklisted by the system
appattempt_1561367690309_0044_000001	Thu Jul 4 15:33:40 +0800 2019	https://node-ana-coreZxDP:8044	Logs	0	0

Showing 1 to 1 of 1 entries

----End

5.3.3 Using a Flume Client

Scenario

You can use Flume to import collected log information to Kafka.

Prerequisites

- A streaming cluster that contains components such as Flume and Kafka and has Kerberos authentication enabled has been created.

- The streaming cluster can properly communicate with the node where logs are generated.

Using the Flume Client (Versions Earlier Than MRS 3.x)

NOTE

You do not need to perform **Step 2** to **Step 6** for a normal cluster.

Step 1 Install the Flume client.

Install the Flume client in a directory, for example, **/opt/Flumeclient**, on the node where logs are generated by referring to **Installing the Flume Client on Clusters of Versions Earlier Than MRS 3.x**. The Flume client installation directories in the following steps are only examples. Change them to the actual installation directories.

Step 2 Copy the configuration file of the authentication server from the Master1 node to the *Flume client installation directory*/**fusioninsight-flume-***Flume component version number*/**conf** directory on the node where the Flume client is installed.

${BIGDATA_HOME}/MRS_Current/1_*X***_KerberosClient/etc/kdc.conf** is used as the full file path.

In the preceding paths, **X** indicates a random number. Change it based on the site requirements. The file must be saved by the user who installs the Flume client, for example, user **root**.

Step 3 Check the service IP address of any node where the Flume role is deployed.

- For versions earlier than MRS 2.0.1, log in to MRS Manager. Choose **Cluster** > **Services** > **Flume** > **Instance**. Query **Service IP Address** of any node on which the Flume role is deployed.
- For MRS 2.0.1 to versions earlier than 3.x, click the cluster name on the MRS console and choose *Name of the desired cluster* > **Components** > **Flume** > **Instances** to view **Business IP Address** of any node where the Flume role is deployed.

Step 4 Copy the user authentication file from this node to the *Flume client installation directory*/**fusioninsight-flume-Flume component version number/conf** directory on the Flume client node.

${BIGDATA_HOME}/MRS_*XXX***/install/FusionInsight-Flume-***Flume component version number*/**flume/conf/flume.keytab** is used as the full file path.

In the preceding paths, **XXX** indicates the product version number. Change it based on the site requirements. The file must be saved by the user who installs the Flume client, for example, user **root**.

Step 5 Copy the **jaas.conf** file from this node to the **conf** directory on the Flume client node.

${BIGDATA_HOME}/MRS_Current/1_*X***_Flume/etc/jaas.conf** is used as the full file path.

In the preceding path, **X** indicates a random number. Change it based on the site requirements. The file must be saved by the user who installs the Flume client, for example, user **root**.

Step 6 Log in to the Flume client node and go to the client installation directory. Run the following command to modify the file:

vi conf/jaas.conf

Change the full path of the user authentication file defined by **keyTab** to the **Flume client installation directory/fusioninsight-flume-*Flume component version number*/conf** saved in **Step 4**, and save the modification and exit.

Step 7 Run the following command to modify the **flume-env.sh** configuration file of the Flume client:

vi *Flume client installation directory***/fusioninsight-flume-***Flume component version number***/conf/flume-env.sh**

Add the following information after **-XX:+UseCMSCompactAtFullCollection**:

```
-Djava.security.krb5.conf=Flume client installation directory/fusioninsight-flume-1.9.0/conf/kdc.conf -
Djava.security.auth.login.config=Flume client installation directory/fusioninsight-flume-1.9.0/conf/jaas.conf -
Dzookeeper.request.timeout=120000
```

Example: **"-XX:+UseCMSCompactAtFullCollection -Djava.security.krb5.conf=/opt/FlumeClient/fusioninsight-flume-***Flume component version number***/conf/kdc.conf -Djava.security.auth.login.config=/opt/FlumeClient/fusioninsight-flume-***Flume component version number***/conf/jaas.conf -Dzookeeper.request.timeout=120000"**

Change *Flume client installation directory* to the actual installation directory. Then save and exit.

Step 8 Run the following command to restart the Flume client:

cd *Flume client installation directory***/fusioninsight-flume-***Flume component version number***/bin**

./flume-manage.sh restart

Example:

cd /opt/FlumeClient/fusioninsight-flume-*Flume component version number***/bin**

./flume-manage.sh restart

Step 9 Run the following command to configure and save jobs in the Flume client configuration file **properties.properties** based on service requirements.

vi *Flume client installation directory***/fusioninsight-flume-***Flume component version number***/conf/properties.properties**

The following uses SpoolDir Source+File Channel+Kafka Sink as an example:

```
##########################################################################
###############
client.sources = static_log_source
client.channels = static_log_channel
client.sinks = kafka_sink
##########################################################################
###############
#LOG_TO_HDFS_ONLINE_1

client.sources.static_log_source.type = spooldir
client.sources.static_log_source.spoolDir = Monitoring directory
client.sources.static_log_source.fileSuffix = .COMPLETED
client.sources.static_log_source.ignorePattern = ^$
client.sources.static_log_source.trackerDir = Metadata storage path during transmission
client.sources.static_log_source.maxBlobLength = 16384
client.sources.static_log_source.batchSize = 51200
client.sources.static_log_source.inputCharset = UTF-8
client.sources.static_log_source.deserializer = LINE
client.sources.static_log_source.selector.type = replicating
client.sources.static_log_source.fileHeaderKey = file
client.sources.static_log_source.fileHeader = false
client.sources.static_log_source.basenameHeader = true
client.sources.static_log_source.basenameHeaderKey = basename
client.sources.static_log_source.deletePolicy = never

client.channels.static_log_channel.type = file
client.channels.static_log_channel.dataDirs = Data cache path. Multiple paths, separated by commas (,), can
be configured to improve performance.
client.channels.static_log_channel.checkpointDir = Checkpoint storage path
client.channels.static_log_channel.maxFileSize = 2146435071
client.channels.static_log_channel.capacity = 1000000
client.channels.static_log_channel.transactionCapacity = 612000
client.channels.static_log_channel.minimumRequiredSpace = 524288000

client.sinks.kafka_sink.type = org.apache.flume.sink.kafka.KafkaSink
client.sinks.kafka_sink.kafka.topic = Topic to which data is written, for example, flume_test
client.sinks.kafka_sink.kafka.bootstrap.servers = XXX.XXX.XXX.XXX:Kafka port number,XXX.XXX.XXX.XXX:Kafka
port number,XXX.XXX.XXX.XXX:Kafka port number
client.sinks.kafka_sink.flumeBatchSize = 1000
client.sinks.kafka_sink.kafka.producer.type = sync
client.sinks.kafka_sink.kafka.security.protocol = SASL_PLAINTEXT
client.sinks.kafka_sink.kafka.kerberos.domain.name = Kafka domain name. This parameter is mandatory for
a security cluster, for example, hadoop.xxx.com.
client.sinks.kafka_sink.requiredAcks = 0

client.sources.static_log_source.channels = static_log_channel
client.sinks.kafka_sink.channel = static_log_channel
```

NOTE

- **client.sinks.kafka_sink.kafka.topic**: Topic to which data is written. If the topic does not exist in Kafka, it is automatically created by default.
- **client.sinks.kafka_sink.kafka.bootstrap.servers**: List of Kafka Brokers, which are separated by commas (,). By default, the port is **21007** for a security cluster and **9092** for a normal cluster.
- **client.sinks.kafka_sink.kafka.security.protocol**: The value is **SASL_PLAINTEXT** for a security cluster and **PLAINTEXT** for a normal cluster.
- **client.sinks.kafka_sink.kafka.kerberos.domain.name**:

  You do not need to set this parameter for a normal cluster. For a security cluster, the value of this parameter is the value of **kerberos.domain.name** in the Kafka cluster.

  Obtain the value by checking **${BIGDATA_HOME}/MRS_Current/1_*X*_Broker/etc/server.properties** on the node where the broker instance resides.

  In the preceding paths, **X** indicates a random number. Change it based on site requirements. The file must be saved by the user who installs the Flume client, for example, user **root**.

Step 10 After the parameters are set and saved, the Flume client automatically loads the content configured in **properties.properties**. When new log files are generated by spoolDir, the files are sent to Kafka producers and can be consumed by Kafka consumers.

----End

Using the Flume Client (MRS 3.x or Later)

NOTE

You do not need to perform **Step 2** to **Step 6** for a normal cluster.

Step 1 Install the Flume client.

Install the Flume client in a directory, for example, **/opt/Flumeclient**, on the node where logs are generated by referring to **Installing the Flume Client on MRS 3.x or Later Clusters**. The Flume client installation directories in the following steps are only examples. Change them to the actual installation directories.

Step 2 Copy the configuration file of the authentication server from the Master1 node to the *Flume client installation directory***/fusioninsight-flume-***Flume component version number***/conf** directory on the node where the Flume client is installed.

The full file path is **${BIGDATA_HOME}/FusionInsight_BASE_***XXX***/1_***X***_KerberosClient/etc/kdc.conf**. In the preceding path, **XXX** indicates the product version number. **X** indicates a random number. Replace them based on site requirements. The file must be saved by the user who installs the Flume client, for example, user **root**.

Step 3 Check the service IP address of any node where the Flume role is deployed.

Log in to FusionInsight Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**. Choose **Cluster** > **Services** > **Flume** > **Instance**. Check the service IP address of any node where the Flume role is deployed.

Step 4 Copy the user authentication file from this node to the *Flume client installation directory***/fusioninsight-flume-Flume component version number/conf** directory on the Flume client node.

The full file path is **${BIGDATA_HOME}/FusionInsight_Porter_***XXX***/install/FusionInsight-Flume-***Flume component version number***/flume/conf/flume.keytab**.

In the preceding paths, **XXX** indicates the product version number. Change it based on the site requirements. The file must be saved by the user who installs the Flume client, for example, user **root**.

Step 5 Copy the **jaas.conf** file from this node to the **conf** directory on the Flume client node.

The full file path is **${BIGDATA_HOME}/FusionInsight_Current/1_***X***_Flume/etc/jaas.conf**.

In the preceding path, **X** indicates a random number. Change it based on the site requirements. The file must be saved by the user who installs the Flume client, for example, user **root**.

Step 6 Log in to the Flume client node and go to the client installation directory. Run the following command to modify the file:

vi conf/jaas.conf

Change the full path of the user authentication file defined by **keyTab** to the **Flume client installation directory/fusioninsight-flume-*Flume component version number*/conf** saved in **Step 4**, and save the modification and exit.

Step 7 Run the following command to modify the **flume-env.sh** configuration file of the Flume client:

vi *Flume client installation directory***/fusioninsight-flume-***Flume component version number***/conf/flume-env.sh**

Add the following information after **-XX:+UseCMSCompactAtFullCollection**:

```
-Djava.security.krb5.conf=Flume client installation directory/fusioninsight-flume-1.9.0/conf/kdc.conf -
Djava.security.auth.login.config=Flume client installation directory/fusioninsight-flume-1.9.0/conf/jaas.conf -
Dzookeeper.request.timeout=120000
```

Example: **"-XX:+UseCMSCompactAtFullCollection -Djava.security.krb5.conf=/opt/FlumeClient/fusioninsight-flume-***Flume component version number***/conf/kdc.conf -Djava.security.auth.login.config=/opt/FlumeClient/fusioninsight-flume-***Flume component version number***/conf/jaas.conf -Dzookeeper.request.timeout=120000"**

Change *Flume client installation directory* to the actual installation directory. Then save and exit.

Step 8 Run the following command to restart the Flume client:

cd *Flume client installation directory***/fusioninsight-flume-***Flume component version number***/bin**

./flume-manage.sh restart

Example:

cd /opt/FlumeClient/fusioninsight-flume-*Flume component version number***/bin**

./flume-manage.sh restart

Step 9 Configure jobs based on actual service scenarios.

- Some parameters, for MRS 3.*x* or later, can be configured on Manager.
- Set the parameters in the **properties.properties** file. The following uses SpoolDir Source+File Channel+Kafka Sink as an example.

  Run the following command on the node where the Flume client is installed. Configure and save jobs in the Flume client configuration file **properties.properties** based on actual service requirements.

  vi *Flume client installation directory***/fusioninsight-flume-***Flume component version number***/conf/properties.properties**

```
#########################################################################
###################
client.sources = static_log_source
client.channels = static_log_channel
client.sinks = kafka_sink
#########################################################################
###################
#LOG_TO_HDFS_ONLINE_1

client.sources.static_log_source.type = spooldir
client.sources.static_log_source.spoolDir = Monitoring directory
client.sources.static_log_source.fileSuffix = .COMPLETED
client.sources.static_log_source.ignorePattern = ^$
client.sources.static_log_source.trackerDir = Metadata storage path during transmission
client.sources.static_log_source.maxBlobLength = 16384
client.sources.static_log_source.batchSize = 51200
client.sources.static_log_source.inputCharset = UTF-8
client.sources.static_log_source.deserializer = LINE
client.sources.static_log_source.selector.type = replicating
client.sources.static_log_source.fileHeaderKey = file
client.sources.static_log_source.fileHeader = false
client.sources.static_log_source.basenameHeader = true
client.sources.static_log_source.basenameHeaderKey = basename
client.sources.static_log_source.deletePolicy = never

client.channels.static_log_channel.type = file
client.channels.static_log_channel.dataDirs = Data cache path. Multiple paths, separated by commas
(,), can be configured to improve performance.
client.channels.static_log_channel.checkpointDir = Checkpoint storage path
client.channels.static_log_channel.maxFileSize = 2146435071
client.channels.static_log_channel.capacity = 1000000
client.channels.static_log_channel.transactionCapacity = 612000
client.channels.static_log_channel.minimumRequiredSpace = 524288000

client.sinks.kafka_sink.type = org.apache.flume.sink.kafka.KafkaSink
client.sinks.kafka_sink.kafka.topic = Topic to which data is written, for example, flume_test
client.sinks.kafka_sink.kafka.bootstrap.servers = XXX.XXX.XXX.XXX:Kafka port
number,XXX.XXX.XXX.XXX:Kafka port number,XXX.XXX.XXX.XXX:Kafka port number
client.sinks.kafka_sink.flumeBatchSize = 1000
client.sinks.kafka_sink.kafka.producer.type = sync
client.sinks.kafka_sink.kafka.security.protocol = SASL_PLAINTEXT
client.sinks.kafka_sink.kafka.kerberos.domain.name = Kafka domain name. This parameter is
mandatory for a security cluster, for example, hadoop.xxx.com.
client.sinks.kafka_sink.requiredAcks = 0

client.sources.static_log_source.channels = static_log_channel
client.sinks.kafka_sink.channel = static_log_channel
```

NOTE

- **client.sinks.kafka_sink.kafka.topic**: Topic to which data is written. If the topic does not exist in Kafka, it is automatically created by default.
- **client.sinks.kafka_sink.kafka.bootstrap.servers**: List of Kafka Brokers, which are separated by commas (,). By default, the port is **21007** for a security cluster and **9092** for a normal cluster.
- **client.sinks.kafka_sink.kafka.security.protocol**: The value is **SASL_PLAINTEXT** for a security cluster and **PLAINTEXT** for a normal cluster.
- **client.sinks.kafka_sink.kafka.kerberos.domain.name**:

  You do not need to set this parameter for a normal cluster. For a security cluster, the value of this parameter is the value of **kerberos.domain.name** in the Kafka cluster.

  Obtain the value by checking **${BIGDATA_HOME}/MRS_Current/1_*X*_Broker/etc/server.properties** on the node where the broker instance resides.

  In the preceding paths, **X** indicates a random number. Change it based on site requirements. The file must be saved by the user who installs the Flume client, for example, user **root**.

Step 10 After the parameters are set and saved, the Flume client automatically loads the content configured in **properties.properties**. When new log files are generated by spoolDir, the files are sent to Kafka producers and can be consumed by Kafka consumers.

----End

5.3.4 Using an HBase Client

Scenario

This section describes how to use the HBase client in an O&M scenario or a service scenario.

Prerequisites

- The client has been installed. For example, the installation directory is **/opt/hadoopclient**. The client directory in the following operations is only an example. Change it to the actual installation directory.
- Service component users are created by the administrator as required.

  A machine-machine user needs to download the **keytab** file and a human-machine user needs to change the password upon the first login.
- If a non-**root** user uses the HBase client, ensure that the owner of the HBase client directory is this user. Otherwise, run the following command to change the owner.

  chown user:group -R *Client installation directory***/HBase**

Using the HBase Client (Versions Earlier Than MRS 3.x)

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client directory:

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the current user. The current user must have the permission to create HBase tables. For details about how to configure a role with corresponding permissions, see To bind a role to a user, see . If Kerberos authentication is disabled for the current cluster, skip this step.

kinit *Component service user*

For example, **kinit hbaseuser**.

Step 5 Run the following HBase client command:

hbase shell

----End

Using the HBase Client (MRS 3.x or Later)

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client directory:

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 If you use the client to connect to a specific HBase instance in a scenario where multiple HBase instances are installed, run the following command to load the environment variables of the instance. Otherwise, skip this step. For example, to load the environment variables of the HBase2 instance, run the following command:

source HBase2/component_env

Step 5 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the current user. The current user must have the permission to create HBase tables. For details about how to configure a role with corresponding permissions, see To bind a role to a user, see . If Kerberos authentication is disabled for the current cluster, skip this step.

kinit *Component service user*

For example, **kinit hbaseuser**.

Step 6 Run the following HBase client command:

hbase shell

----End

Common HBase client commands

The following table lists common HBase client commands. For more commands, see **http://hbase.apache.org/2.2/book.html**.

Table 5-4 HBase client commands

Command	Description
create	Used to create a table, for example, **create 'test', 'f1', 'f2', 'f3'**.
disable	Used to disable a specified table, for example, **disable 'test'**.
enable	Used to enable a specified table, for example, **enable 'test'**.
alter	Used to alter the table structure. You can run the **alter** command to add, modify, or delete column family information and table-related parameter values, for example, **alter 'test', {NAME => 'f3', METHOD => 'delete'}**.
describe	Used to obtain the table description, for example, **describe 'test'**.
drop	Used to delete a specified table, for example, **drop 'test'**. Before deleting a table, you must stop it.
put	Used to write the value of a specified cell, for example, **put 'test','r1','f1:c1','myvalue1'**. The cell location is unique and determined by the table, row, and column.

Command	Description
get	Used to get the value of a row or the value of a specified cell in a row, for example, **get 'test','r1'**.
scan	Used to query table data, for example, **scan 'test'**. The table name and scanner must be specified in the command.

5.3.5 Using an HDFS Client

Scenario

This section describes how to use the HDFS client in an O&M scenario or service scenario.

Prerequisites

- The client has been installed.

  For example, the installation directory is **/opt/hadoopclient**. The client directory in the following operations is only an example. Change it to the actual installation directory.
- Service component users are created by the administrator as required. In security mode, machine-machine users need to download the keytab file. A human-machine user needs to change the password upon the first login. (This operation is not required in normal mode.)

Using the HDFS Client

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client installation directory:

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 If the cluster is in security mode, run the following command to authenticate the user. In normal mode, user authentication is not required.

kinit *Component service user*

Step 5 Run the HDFS Shell command. Example:

hdfs dfs -ls /

----End

Common HDFS Client Commands

The following table lists common HDFS client commands.

For more commands, see **https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/CommandsManual.html#User_Commands**.

Table 5-5 Common HDFS client commands

Command	Description	Example
hdfs dfs -mkdir *Folder name*	Used to create a folder.	**hdfs dfs -mkdir /tmp/mydir**
hdfs dfs -ls *Folder name*	Used to view a folder.	**hdfs dfs -ls /tmp**
hdfs dfs -put *Local file on the client node Specified HDFS path*	Used to upload a local file to a specified HDFS path.	**hdfs dfs -put /opt/test.txt /tmp** Upload the **/opt/test.txt** file on the client node to the **/tmp** directory of HDFS.
hdfs dfs -get *Specified file on HDFS Specified path on the client node*	Used to download the HDFS file to the specified local path.	**hdfs dfs -get /tmp/test.txt /opt/** Download the **/tmp/test.txt** file on HDFS to the **/opt** path on the client node.
hdfs dfs -rm -r -f *Specified folder on HDFS*	Used to delete a folder.	**hdfs dfs -rm -r -f /tmp/mydir**
hdfs dfs -chmod *Permission parameter File directory*	Used to configure the HDFS directory permission for a user.	**hdfs dfs -chmod 700 /tmp/test**

Client-related FAQs

1. What do I do when the HDFS client exits abnormally and error message "java.lang.OutOfMemoryError" is displayed after the HDFS client command is running?

   This problem occurs because the memory required for running the HDFS client exceeds the preset upper limit (128 MB by default). You can change the memory upper limit of the client by modifying **CLIENT_GC_OPTS** in *<Client installation path>***/HDFS/component_env**. For example, if you want to set the upper limit to 1 GB, run the following command:

   ```
   CLIENT_GC_OPTS="-Xmx1G"
   ```

   After the modification, run the following command to make the modification take effect:

   source <*Client installation path*>/**/bigdata_env**

2. How do I set the log level when the HDFS client is running?

   By default, the logs generated during the running of the HDFS client are printed to the console. The default log level is INFO. To enable the DEBUG log level for fault locating, run the following command to export an environment variable:

export HADOOP_ROOT_LOGGER=DEBUG,console

Then run the HDFS Shell command to generate the DEBUG logs.

If you want to print INFO logs again, run the following command:

export HADOOP_ROOT_LOGGER=INFO,console

3. How do I delete HDFS files permanently?

   HDFS provides a recycle bin mechanism. Typically, after an HDFS file is deleted, the file is moved to the recycle bin of HDFS. If the file is no longer needed and the storage space needs to be released, clear the corresponding recycle bin directory, for example, **hdfs://hacluster/user/xxx/.Trash/Current/***xxx*.

5.3.6 Using a Hive Client

Scenario

This section guides users to use a Hive client in an O&M or service scenario.

Prerequisites

- The client has been installed. For example, the client is installed in the **/opt/hadoopclient** directory. The client directory in the following operations is only an example. Change it to the actual installation directory.
- Service component users are created by the administrator as required. In security mode, machine-machine users need to download the keytab file. A human-machine user must change the password upon the first login.

Using the Hive Client (Versions Earlier Than MRS 3.x)

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client installation directory:

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 Log in to the Hive client based on the cluster authentication mode.

- In security mode, run the following command to complete user authentication and log in to the Hive client:

  kinit *Component service user*

  beeline

- In common mode, run the following command to log in to the Hive client. If no component service user is specified, the current OS user is used to log in to the Hive client.

  beeline -n *component service user*

NOTE

After a beeline connection is established, you can compile and submit HQL statements to execute related tasks. To run the Catalog client command, you need to run the **!q** command first to exit the beeline environment.

Step 5 Run the following command to execute the HCatalog client command:

hcat -e *"cmd"*

cmd must be a Hive DDL statement, for example, **hcat -e "show tables"**.

NOTE

- To use the HCatalog client, choose **More** > **Download Client** on the service page to download the clients of all services. This restriction does not apply to the beeline client.
- Due to permission model incompatibility, tables created using the HCatalog client cannot be accessed on the HiveServer client. However, the tables can be accessed on the WebHCat client.
- If you use the HCatalog client in Normal mode, the system performs DDL commands using the current user who has logged in to the operating system.
- Exit the beeline client by running the **!q** command instead of by pressing **Ctrl + c**. Otherwise, the temporary files generated by the connection cannot be deleted and a large number of junk files will be generated as a result.
- If multiple statements need to be entered during the use of beeline clients, separate the statements from each other using semicolons (**;**) and set the value of **entireLineAsCommand** to **false**.

  Setting method: If beeline has not been started, run the **beeline --entireLineAsCommand=false** command. If the beeline has been started, run the **!set entireLineAsCommand false** command.

  After the setting, if a statement contains semicolons (**;**) that do not indicate the end of the statement, escape characters must be added, for example, **select concat_ws('\;', collect_set(col1)) from tbl**.

----End

Using the Hive Client (MRS 3.x or Later)

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client installation directory:

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 MRS 3.*X* supports multiple Hive instances. If you use the client to connect to a specific Hive instance in a scenario when multiple Hive instances are installed, run the following command to load the environment variables of the instance. Otherwise, skip this step. For example, load the environment variables of the Hive2 instance.

source Hive2/component_env

Step 5 Log in to the Hive client based on the cluster authentication mode.

- In security mode, run the following command to complete user authentication and log in to the Hive client:

  kinit *Component service user*

  beeline

- In common mode, run the following command to log in to the Hive client. If no component service user is specified, the current OS user is used to log in to the Hive client.

  beeline -n *component service user*

Step 6 Run the following command to execute the HCatalog client command:

hcat -e *"cmd"*

cmd must be a Hive DDL statement, for example, **hcat -e "show tables"**.

NOTE

- To use the HCatalog client, choose **More** > **Download Client** on the service page to download the clients of all services. This restriction does not apply to the beeline client.
- Due to permission model incompatibility, tables created using the HCatalog client cannot be accessed on the HiveServer client. However, the tables can be accessed on the WebHCat client.
- If you use the HCatalog client in Normal mode, the system performs DDL commands using the current user who has logged in to the operating system.
- Exit the beeline client by running the **!q** command instead of by pressing **Ctrl + C**. Otherwise, the temporary files generated by the connection cannot be deleted and a large number of junk files will be generated as a result.
- If multiple statements need to be entered during the use of beeline clients, separate the statements from each other using semicolons (**;**) and set the value of **entireLineAsCommand** to **false**.

  Setting method: If beeline has not been started, run the **beeline --entireLineAsCommand=false** command. If the beeline has been started, run the **!set entireLineAsCommand false** command.

  After the setting, if a statement contains semicolons (**;**) that do not indicate the end of the statement, escape characters must be added, for example, **select concat_ws('\;', collect_set(col1)) from tbl**.

----End

Common Hive Client Commands

The following table lists common Hive Beeline commands.

For more commands, see **https://cwiki.apache.org/confluence/display/Hive/HiveServer2+Clients#HiveServer2Clients-BeelineCommands**.

Table 5-6 Common Hive Beeline commands

Command	Description
set <key>=<value>	Sets the value of a specific configuration variable (key). **NOTE** If the variable name is incorrectly spelled, the Beeline does not display an error.
set	Prints the list of configuration variables overwritten by users or Hive.

Command	Description
set -v	Prints all configuration variables of Hadoop and Hive.
add FILE[S] <filepath> <filepath>*add JAR[S] <filepath> <filepath>*add ARCHIVE[S] <filepath> <filepath>*	Adds one or more files, JAR files, or ARCHIVE files to the resource list of the distributed cache.
add FILE[S] <ivyurl> <ivyurl>* add JAR[S] <ivyurl> <ivyurl>* add ARCHIVE[S] <ivyurl> <ivyurl>*	Adds one or more files, JAR files, or ARCHIVE files to the resource list of the distributed cache using the lvy URL in the **ivy://goup:module:version?query_string** format.
list FILE[S]list JAR[S]list ARCHIVE[S]	Lists the resources that have been added to the distributed cache.
list FILE[S] <filepath>*list JAR[S] <filepath>*list ARCHIVE[S] <filepath>*	Checks whether given resources have been added to the distributed cache.
delete FILE[S] <filepath>*delete JAR[S] <filepath>*delete ARCHIVE[S] <filepath>*	Deletes resources from the distributed cache.
delete FILE[S] <ivyurl> <ivyurl>* delete JAR[S] <ivyurl> <ivyurl>* delete ARCHIVE[S] <ivyurl> <ivyurl>*	Delete the resource added using **<ivyurl>** from the distributed cache.
reload	Enable HiveServer2 to discover the change of the JAR file **hive.reloadable.aux.jars.path** in the specified path. (You do not need to restart HiveServer2.) Change actions include adding, deleting, or updating JAR files.
dfs <dfs command>	Runs the **dfs** command.
<query string>	Executes the Hive query and prints the result to the standard output.

5.3.7 Using an Impala Client

Impala is a massively parallel processing (MPP) SQL query engine for processing vast amounts of data stored in Hadoop clusters. It is an open source software written in C++ and Java. It provides high performance and low latency compared with other SQL engines for Hadoop.

Background

Suppose a user develops an application to manage users who use service A in an enterprise. The procedure of operating service A on the Impala client is as follows:

Operations on common tables:

- Create the **user_info** table.
- Add users' educational backgrounds and titles to the table.
- Query user names and addresses by user ID.
- Delete the user information table after service A ends.

Table 5-7 User information

No.	Name	Gender	Age	Address
12005000201	A	Male	19	City A
12005000202	B	Female	23	City B
12005000203	C	Male	26	City C
12005000204	D	Male	18	City D
12005000205	E	Female	21	City E
12005000206	F	Male	32	City F
12005000207	G	Female	29	City G
12005000208	H	Female	30	City H
12005000209	I	Male	26	City I
12005000210	J	Female	25	City J

Prerequisites

The client has been installed. For example, the client is installed in the **/opt/hadoopclient** directory. The client directory in the following operations is only an example. Change it to the actual installation directory.

Procedure

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client installation directory:

```
cd /opt/hadoopclient
```

Step 3 Run the following command to configure environment variables:

```
source bigdata_env
```

Step 4 Run the Impala client command to implement service A.

Run the client command of the Impala component directly.

impala-shell

NOTE

By default, **impala-shell** attempts to connect to the Impala daemon on port 21000 of **localhost**. To connect to another host, use the **-i < host:port >** option, for example, **impala-shell -i** *xxx.xxx.xxx.xxx***:21000**. To automatically connect to a specific Impala database, use the **-d <database>** option. For example, if all your Kudu tables are in the **impala_kudu** database, **-d impala_kudu** can use this database. To exit the Impala Shell, run the **quit** command.

Operations on internal tables:

1. Create the **user_info** user information table according to **Table 5-7** and add data to it.

   ```
   create table user_info(id string,name string,gender string,age int,addr string);
   insert into table user_info(id,name,gender,age,addr) values("12005000201", "A", "Male", 19, "City A");
   ```

   ... (Other statements are the same.)

2. Add users' educational backgrounds and titles to the **user_info** table.

   For example, to add educational background and title information about user 12005000201, run the following commands.

   ```
   alter table user_info add columns(education string,technical string);
   ```

3. Query user names and addresses by user ID.

   For example, to query the name and address of user 12005000201, run the following command:

   ```
   select name,addr from user_info where id='12005000201';
   ```

4. Delete the user information table:

   ```
   drop table user_info;
   ```

Operations on external partition tables:

Create an external partition table and import data.

1. Create a path for storing external table data.
   - Security mode (Kerberos authentication is enabled for clusters)

     cd /opt/hadoopclient

     source bigdata_env

     kinit *hive*

     NOTE

     The user must have the hive administrator permissions.

     impala-shell

     hdfs dfs -mkdir /hive

     hdfs dfs -mkdir /hive/user_info
   - Normal mode (Kerberos authentication is disabled for clusters)

     su - omm

     cd /opt/hadoopclient

     source bigdata_env

impala-shell

hdfs dfs -mkdir /hive

hdfs dfs -mkdir /hive/user_info

2. Create a table.

```
create external table user_info(id string,name string,gender string,age int,addr string) partitioned
by(year string) row format delimited fields terminated by ' ' lines terminated by '\n' stored as textfile
location '/hive/user_info';
```

NOTE

fields terminated indicates delimiters, for example, spaces.

lines terminated indicates line breaks, for example, **\n**.

/hive/user_info indicates the path of the data file.

3. Import data.

   a. Execute the **insert** statement to insert data.

   ```
   insert into user_info partition(year="2018") values ("12005000201", "A", "Male", 19, "City A");
   ```

   b. Run the **load data** command to import file data.

      i. Create a file based on the data in **Table 5-7**. For example, the file name is **txt.log**. Fields are separated by space, and the line feed characters are used as the line breaks.

      ii. Upload the file to HDFS.

      hdfs dfs -put txt.log /tmp

      iii. Load data to the table.

      load data inpath '/tmp/txt.log' into table user_info partition (year='2018');

4. Query the imported data:

```
select * from user_info;
```

5. Delete the user information table:

```
drop table user_info;
```

----End

5.3.8 Using a Kafka Client

Scenario

You can create, query, and delete topics on a cluster client.

Prerequisites

The client has been installed. For example, the client is installed in the **/opt/hadoopclient** directory. The client directory in the following operations is only an example. Change it to the actual installation directory.

Using the Kafka Client (Versions Earlier Than MRS 3.x)

Step 1 Access the ZooKeeper instance page.

- For versions earlier than MRS 2.0.1, log in to MRS Manager and choose **Services** > **ZooKeeper** > **Instance**.

- For MRS 2.0.1 or later to versions earlier than 3.x, click the cluster name on the MRS console and choose **Components** > **ZooKeeper** > **Instances**.

  NOTE

  If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

Step 2 View the IP addresses of the ZooKeeper role instance.

Record any IP address of the ZooKeeper instance.

Step 3 Log in to the node where the client is installed.

Step 4 Run the following command to switch to the client directory, for example, **/opt/hadoopclient/Kafka/kafka/bin**.

cd /opt/hadoopclient/Kafka/kafka/bin

Step 5 Run the following command to configure environment variables:

source /opt/hadoopclient/bigdata_env

Step 6 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the current user. If Kerberos authentication is disabled for the current cluster, skip this step.

kinit *Kafka user*

Step 7 Create a topic.

sh kafka-topics.sh --create --topic *Topic name* **--partitions** *Number of partitions occupied by the topic* **--replication-factor** *Number of replicas of the topic* **--zookeeper** *IP address of the node where the ZooKeeper instance resides:clientPort***/kafka**

Example: **sh kafka-topics.sh --create --topic TopicTest --partitions 3 --replication-factor 3 --zookeeper 10.10.10.100:2181/kafka**

Step 8 Run the following command to view the topic information in the cluster:

sh kafka-topics.sh --list --zookeeper *IP address of the node where the ZooKeeper instance resides:clientPort***/kafka**

Example: **sh kafka-topics.sh --list --zookeeper 10.10.10.100:2181/kafka**

Step 9 Delete the topic created in **Step 7**.

sh kafka-topics.sh --delete --topic *Topic name* **--zookeeper** *IP address of the node where the ZooKeeper instance resides.clientPort***/kafka**

Example: **sh kafka-topics.sh --delete --topic TopicTest --zookeeper 10.10.10.100:2181/kafka**

Type **y** and press **Enter**.

----End

Using the Kafka Client (MRS 3.x or Later)

Step 1 Access the ZooKeeper instance page.

Log in to FusionInsight Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**. Choose **Cluster** > *Name of the desired cluster* > **Services** > **ZooKeeper** > **Instance**.

Step 2 View the IP addresses of the ZooKeeper role instance.

Record any IP address of the ZooKeeper instance.

Step 3 Log in to the node where the client is installed.

Step 4 Run the following command to switch to the client directory, for example, **/opt/hadoopclient/Kafka/kafka/bin**.

cd /opt/hadoopclient/Kafka/kafka/bin

Step 5 Run the following command to configure environment variables:

source /opt/hadoopclient/bigdata_env

Step 6 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the current user. If Kerberos authentication is disabled for the current cluster, skip this step.

kinit *Kafka user*

Step 7 Log in to FusionInsight Manager, choose **Cluster** > **Name of the desired cluster** > **Services** > **ZooKeeper**, and click the **Configurations** tab and then **All Configurations**. On the displayed page, search for the **clientPort** parameter and record its value.

Step 8 Create a topic.

sh kafka-topics.sh --create --topic *Topic name* **--partitions** *Number of partitions occupied by the topic* **--replication-factor** *Number of replicas of the topic* **--zookeeper** *IP address of the node where the ZooKeeper instance resides:clientPort***/kafka**

Example: **sh kafka-topics.sh --create --topic TopicTest --partitions 3 --replication-factor 3 --zookeeper 10.10.10.100:2181/kafka**

Step 9 Run the following command to view the topic information in the cluster:

sh kafka-topics.sh --list --zookeeper *IP address of the node where the ZooKeeper instance resides:clientPort***/kafka**

Example: **sh kafka-topics.sh --list --zookeeper 10.10.10.100:2181/kafka**

Step 10 Delete the topic created in **Step 8**.

sh kafka-topics.sh --delete --topic *Topic name* **--zookeeper** *IP address of the node where the ZooKeeper instance resides.clientPort***/kafka**

Example: **sh kafka-topics.sh --delete --topic TopicTest --zookeeper 10.10.10.100:2181/kafka**

----End

5.3.9 Using a Kudu Client

Kudu is a columnar storage manager developed for the Apache Hadoop platform. Kudu shares the common technical properties of Hadoop ecosystem applications. It is horizontally scalable and supports highly available operations.

Prerequisites

The cluster client has been installed. For example, the client is installed in the **/opt/hadoopclient** directory. The client directory in the following operations is only an example. Change it to the actual installation directory.

Procedure

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client installation directory:

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 Run the Kudu command line tool.

Run the command line tool of the Kudu component to view help information.

kudu -h

The command output is as follows:

```
Usage: kudu <command> [<args>]

<command> can be one of the following:
         cluster   Operate on a Kudu cluster
        diagnose   Diagnostic tools for Kudu servers and clusters
              fs   Operate on a local Kudu filesystem
             hms   Operate on remote Hive Metastores
   local_replica   Operate on local tablet replicas via the local filesystem
          master   Operate on a Kudu Master
             pbc   Operate on PBC (protobuf container) files
            perf   Measure the performance of a Kudu cluster
  remote_replica   Operate on remote tablet replicas on a Kudu Tablet Server
           table   Operate on Kudu tables
          tablet   Operate on remote Kudu tablets
            test   Various test actions
         tserver   Operate on a Kudu Tablet Server
             wal   Operate on WAL (write-ahead log) files
```

NOTE

The Kudu command line tool does not support DDL and DML operations, but provides the refined query function for the **cluster**, **master**, **tserver**, **fs**, and **table** parameters.

Common operations:

- Check the tables in the current cluster.

  kudu table list *KuduMaster instance IP1:7051, KuduMaster instance IP2:7051, KuduMaster instance IP3:7051*
- Query the configurations of the KuduMaster instance of the Kudu service.

kudu master get_flags *KuduMaster instance IP:7051*

- Query the schema of a table.

  kudu table describe *KuduMaster instance IP1:7051, KuduMaster instance IP2:7051, KuduMaster instance IP3:7051 Table name*

- Delete a table.

  kudu table delete *KuduMaster instance IP1:7051, KuduMaster instance IP2:7051, KuduMaster instance IP3:7051 Table name*

  NOTE

  To obtain the IP address of the KuduMaster instance, choose **Components** > **Kudu** > **Instances** on the cluster details page.

----End

5.3.10 Using the Oozie Client

Scenario

This section describes how to use the Oozie client in an O&M scenario or service scenario.

Prerequisites

- The client has been installed. For example, the installation directory is **/opt/client**. The client directory in the following operations is only an example.
- Service component users are created by the administrator as required. In security mode, machine-machine users need to download the keytab file. A human-machine user must change the password upon the first login.

Using the Oozie Client

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to switch to the client installation directory (change it to the actual installation directory):

```
cd /opt/client
```

Step 3 Run the following command to configure environment variables:

```
source bigdata_env
```

Step 4 Check the cluster authentication mode.

- If the cluster is in security mode, run the following command to authenticate the user: *exampleUser* indicates the name of the user who submits tasks.

  kinit *exampleUser*

- If the cluster is in normal mode, go to **Step 5**.

Step 5 Perform the following operations to configure Hue:

1. Configure the Spark2x environment (skip this step if the Spark2x task is not involved):

   ```
   hdfs dfs -put /opt/client/Spark2x/spark/jars/*.jar /user/oozie/share/lib/spark2x/
   ```

When the JAR package in the HDFS directory **/user/oozie/share** changes, you need to restart the Oozie service.

2. Upload the Oozie configuration file and JAR package to HDFS.

   hdfs dfs -mkdir /user/*exampleUser*

   hdfs dfs -put -f /opt/client/Oozie/oozie-client-*/examples /user/*exampleUser***/**

   NOTE

   - *exampleUser* indicates the name of the user who submits tasks.
   - If the user who submits the task and other files except **job.properties** are not changed, client installation directory **Oozie/oozie-client-*/examples** can be repeatedly used after being uploaded to HDFS.
   - Resolve the JAR file conflict between Spark and Yarn about Jetty.

     hdfs dfs -rm -f /user/oozie/share/lib/spark/jetty-all-9.2.22.v20170606.jar
   - In normal mode, if **Permission denied** is displayed during the upload, run the following commands:

     su - omm

     source /opt/client/bigdata_env

     hdfs dfs -chmod -R 777 /user/oozie

     exit

----End

5.3.11 Using a Storm Client

Scenario

This section describes how to use the Storm client in an O&M scenario or service scenario.

Prerequisites

- You have installed the client. For example, the installation directory is **/opt/hadoopclient**.
- Service component users are created by the administrator as required. In security mode, machine-machine users have downloaded the keytab file. A human-machine user must change the password upon the first login. (Not involved in normal mode)

Procedure

Step 1 Prepare the client based on service requirements. Log in to the node where the client is installed.

Log in to the node where the client is installed. For details, see .

Step 2 Run the following command to go to the client installation directory:

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 If multiple Storm instances are installed, run the following command to load the environment variables of a specific instance when running the Storm command to submit the topology. Otherwise, skip this step. The following command uses the instance Storm-2 as an example.

source Storm-2/component_env

Step 5 Run the following command to perform user authentication (skip this step in normal mode):

kinit *Component service user*

Step 6 Run the following command to perform operations on the client:

For example, run the following command:

- **cql**
- **storm**

NOTE

A Storm client cannot be connected to secure and non-secure ZooKeepers at the same time.

----End

5.3.12 Using a Yarn Client

Scenario

This section guides users to use a Yarn client in an O&M or service scenario.

Prerequisites

- The client has been installed.

  For example, the installation directory is **/opt/hadoopclient**. The client directory in the following operations is only an example. Change it to the actual installation directory.
- Service component users are created by the administrator as required. In security mode, machine-machine users need to download the keytab file. A human-machine user must change the password upon the first login. In common mode, you do not need to download the keytab file or change the password.

Using the Yarn Client

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client installation directory:

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 If the cluster is in security mode, run the following command to authenticate the user. In normal mode, user authentication is not required.

kinit *Component service user*

Step 5 Run the Yarn command. The following provides an example:

yarn application -list

----End

Client-related FAQs

1. What Do I Do When the Yarn Client Exits Abnormally and Error Message "java.lang.OutOfMemoryError" Is Displayed After the Yarn Client Command Is Run?

   This problem occurs because the memory required for running the Yarn client exceeds the upper limit (128 MB by default) set on the Yarn client. For clusters of MRS 3.x or later: You can modify **CLIENT_GC_OPTS** in *<Client installation path>***/HDFS/component_env** to change the memory upper limit of the Yarn client. For example, if you want to set the maximum memory to 1 GB, run the following command:

   ```
   export CLIENT_GC_OPTS="-Xmx1G"
   ```

   For clusters earlier than MRS 3.x: You can modify **GC_OPTS_YARN** in *< Client installation path >***/HDFS/component_env** to change the memory upper limit of the Yarn client. For example, if you want to set the maximum memory to 1 GB, run the following command:

   ```
   export GC_OPTS_YARN="-Xmx1G"
   ```

   After the modification, run the following command to make the modification take effect:

   source *<Client installation path>***//bigdata_env**

2. How Can I Set the Log Level When the Yarn Client Is Running?

   By default, the logs generated during the running of the Yarn client are printed to the console. The default log level is INFO. To enable the DEBUG log level for fault locating, run the following command to export an environment variable:

   export YARN_ROOT_LOGGER=DEBUG,console

   Then run the Yarn Shell command to print DEBUG logs.

   If you want to print INFO logs again, run the following command:

   export YARN_ROOT_LOGGER=INFO,console

6 Configuring a Cluster with Storage and Compute Decoupled

6.1 Introduction to Storage-Compute Decoupling

In scenarios that require large storage capacity and elastic compute resources, MRS enables you to store data in OBS and use an MRS cluster for data computing only. In this way, storage and compute are separated.

NOTE

In the big data decoupled storage-compute scenario, the OBS parallel file system must be used to configure a cluster. Using common object buckets will greatly affect the cluster performance.

Process of using the storage-compute decoupling function:

1. Configure a storage-compute decoupled cluster using either of the following methods (agency is recommended):
   - Bind an agency of the ECS type to an MRS cluster to access OBS, preventing the AK/SK from being exposed in the configuration file. For details, see **Configuring a Storage-Compute Decoupled Cluster (Agency)**.
   - Configure the AK/SK in an MRS cluster. The AK/SK will be exposed in the configuration file in plaintext. Exercise caution when performing this operation. For details, see **Configuring a Storage-Compute Decoupled Cluster (AK/SK)**.
2. Use the cluster.

   For details, see the following sections':

   - **Interconnecting Flink with OBS**
   - **Interconnecting Flume with OBS**
   - **Interconnecting HDFS with OBS**
   - **Interconnecting Hive with OBS**
   - **Interconnecting MapReduce with OBS**

- **Interconnecting Spark2x with OBS**
- **Interconnecting Sqoop with External Storage Systems**

6.2 Configuring a Storage-Compute Decoupled Cluster (Agency)

MRS allows you to store data in OBS and use an MRS cluster for data computing only. In this way, storage and compute are separated. You can create an IAM agency, which enables ECS to automatically obtain the temporary AK/SK to access OBS. This prevents the AK/SK from being exposed in the configuration file.

By binding an agency, ECSs or BMSs can manage some of your resources. Determine whether to configure an agency based on the actual service scenario.

MRS provides the following configuration modes for accessing OBS. You can select one of them. The agency mode is recommended.

- Bind an agency of the ECS type to an MRS cluster to access OBS, preventing the AK/SK from being exposed in the configuration file. For details, see the following part in this section.
- Configure the AK/SK in an MRS cluster. The AK/SK will be exposed in the configuration file in plaintext. Exercise caution when performing this operation. For details, see **Configuring a Storage-Compute Decoupled Cluster (AK/SK)**.

This function is available for components Hadoop, Hive, Spark, Presto, and Flink in clusters of .

(Optional) Step 1: Create an ECS Agency with OBS Access Permissions

NOTE

- MRS presets **MRS_ECS_DEFAULT_AGENCY** in the agency list of IAM so that you can select this agency when creating a cluster. This agency has the **OBSOperateAccess** permission and the **CESFullAccess** (only available for users who have enabled fine-grained policies), **CES Administrator**, and **KMS Administrator** permissions in the region where the cluster is located. Do not modify **MRS_ECS_DEFAULT_AGENCY** on IAM.
- If you want to use the preset agency, skip the step for creating an agency. If you want to use a custom agency, perform the following steps to create an agency. (To create or modify an agency, you must have the Security Administrator permission.)

1. Log in to the management console.
2. Choose **Service List** > **Management & Governance** > **Identity and Access Management**.
3. Choose **Agencies**. On the displayed page, click **Create Agency**.
4. Enter an agency name, for example, **mrs_ecs_obs**.
5. Set **Agency Type** to **Cloud service** and select **ECS BMS** to authorize ECS or BMS to invoke OBS.
6. Set **Validity Period** to **Unlimited** and click **Next**.
7. On the displayed page, search for the **OBS OperateAccess** and select it.

8. Click **Next**. On the displayed page, select the desired scope for permissions you selected. By default, **All resources** is selected. Click **Show More** and select **Global resources**.
9. In the dialog box that is displayed, click **OK** to start authorization. After the message "**Authorization successful.**" is displayed, click **Finish**. The agency is successfully created.

Step 2: Create a Cluster with Storage and Compute Separated

You can configure an agency when creating a cluster or bind an agency to an existing cluster to separate storage and compute. This section uses a cluster with Kerberos authentication enabled as an example.

Configuring an agency when creating a cluster:

1. Log in to the MRS management console.
2. Click **Create Cluster**. The page for creating a cluster is displayed.
3. Click the **Custom Config** tab.
4. On the **Custom Config** tab page, set software parameters.
   - **Region**: Select a region as required.
   - **Cluster Name**: You can use the default name. However, you are advised to include a project name abbreviation or date for consolidated memory and easy distinguishing.
   - Cluster Version: Select a cluster version.
   - **Cluster Type**: Select **Analysis cluster** or **Hybrid cluster** and select all components.
   - **Metadata**: Select **Local**.
5. Click **Next** and set hardware parameters.
   - **AZ**: Use the default value.
   - **VPC**: Use the default value.
   - **Subnet**: Use the default value.
   - **Security Group**: Use the default value.
   - **EIP**: Use the default value.
   - **Cluster Node**: Select the number of cluster nodes and node specifications based on site requirements.
6. Click **Next** and set related parameters.
   - **Kerberos Authentication**: This function is enabled by default. You can enable or disable it.
   - **Username**: The default username is **admin**, which is used to log in to MRS Manager.
   - **Password**: Set a password for user **admin**.
   - **Confirm Password**: Enter the password of user **admin** again.
   - **Key Pair**: Select a key pair from the drop-down list to log in to an ECS. Select **"I acknowledge that I have obtained private key file *SSHkey-xxx* and that without this file I will not be able to log in to my ECS.**" If you have never created a key pair, click **View Key Pair** to create or import a key pair. And then, obtain a private key file.

7. In this example, configure an agency and leave other parameters blank. For details about how to configure other parameters, see **(Optional) Advanced Configuration**.

   Agency: Select the agency created in **(Optional) Step 1: Create an ECS Agency with OBS Access Permissions** or **MRS_ECS_DEFAULT_AGENCY** preset in IAM.

8. To enable secure communications, select **Enable**. For details, see **Communication Security Authorization**.
9. and wait until the cluster is created.

   If Kerberos authentication is enabled for a cluster, check whether Kerberos authentication is required. If yes, click **Continue**. If no, click **Back** to disable Kerberos authentication and then create a cluster.

Configuring an agency for an existing cluster:

1. Log in to the MRS management console. In the left navigation pane, choose **Clusters** > **Active Clusters**.
2. Click the name of the cluster to enter its details page.
3. On the **Dashboard** page, click **Synchronize** on the right of **IAM User Sync** to synchronize IAM users.
4. On the **Dashboard** tab page, click **Manage Agency** on the right side of **Agency** to select an agency and click **OK** to bind it. Alternatively, click **Create Agency** to go to the IAM console to create an agency and select it.

Step 3: Create an OBS File System for Storing Data

NOTE

In the big data decoupled storage-compute scenario, the OBS parallel file system must be used to configure a cluster. Using common object buckets will greatly affect the cluster performance.

1. Log in to OBS Console.
2. Choose **Parallel File System** > **Create Parallel File System**.
3. Enter the file system name, for example, **mrs-word001**.

   Set other parameters as required.

4. Click **Create Now**.
5. In the parallel file system list on the OBS console, click the file system name to go to the details page.
6. In the navigation pane, choose **Files** and create the **program** and **input** folders.
   - **program**: Upload the program package to this folder.
   - **input**: Upload the input data to this folder.

Step 4: Accessing the OBS File System

1. Log in to a Master node as user **root**. For details, see **Logging In to an ECS**.
2. Run the following command to set the environment variables:

   For versions earlier than MRS 3.x, run the **source /opt/client/bigdata_env** command.

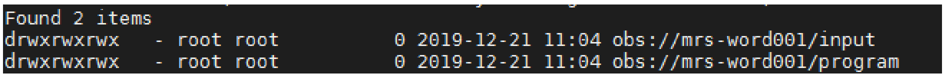
For MRS 3.x or later, run the **source /opt/Bigdata/client/bigdata_env** command.

3. Verify that Hadoop can access OBS.
   a. View the list of files in the file system **mrs-word001**.

      hadoop fs -ls obs://mrs-word001/

   b. Check whether the file list is returned. If it is returned, OBS access is successful.

      Figure 6-1 Returned file list

```
Found 2 items
drwxrwxrwx   - root root          0 2019-12-21 11:04 obs://mrs-word001/input
drwxrwxrwx   - root root          0 2019-12-21 11:04 obs://mrs-word001/program
```

4. Verify that Hive can access OBS.
   a. If Kerberos authentication has been enabled for the cluster, run the following command to authenticate the current user. The current user must have a permission to create Hive tables. For details about how to configure a role with a permission to create Hive tables, see **Creating a Role**. For details about how to create a user and bind a role to the user, see **Creating a User**. If Kerberos authentication is disabled for the current cluster, skip this step.

      kinit ***MRS cluster user***

      Example: **kinit hiveuser**

   b. Run the client command of the Hive component.

      beeline

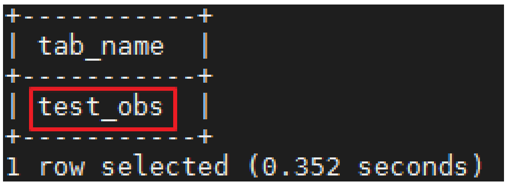
   c. Access the OBS directory in the beeline. For example, run the following command to create a Hive table and specify that data is stored in the **test_obs** directory of the file system **mrs-word001**:

      create table test_obs(a int, b string) row format delimited fields terminated by "," stored as textfile location "obs://mrs-word001/test_obs";

   d. Run the following command to query all tables. If table **test_obs** is displayed in the command output, OBS access is successful.

      show tables;

      Figure 6-2 Returned table name

```
+-----------+
| tab_name  |
+-----------+
| test_obs  |
+-----------+
1 row selected (0.352 seconds)
```

   e. Press **Ctrl+C** to exit the Hive beeline.

5. Verify that Spark can access OBS.
   a. Run the client command of the Spark component.

      spark-beeline

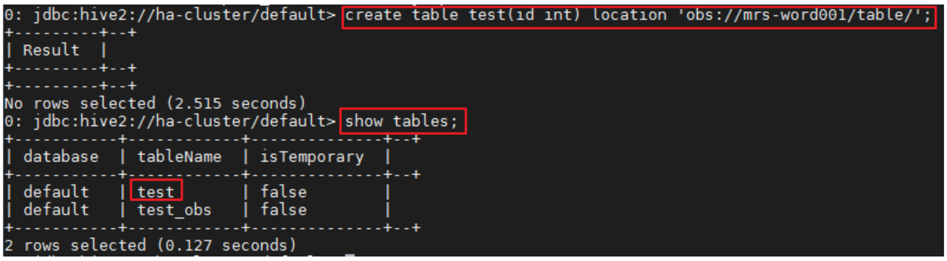
   b. Access OBS in spark-beeline. For example, create table **test** in the **obs://mrs-word001/table/** directory.

create table test(id int) location 'obs://mrs-word001/table/';

c. Run the following command to query all tables. If table **test** is displayed in the command output, OBS access is successful.

show tables;

Figure 6-3 Returned table name



```
0: jdbc:hive2://ha-cluster/default> create table test(id int) location 'obs://mrs-word001/table/';
+---------+--+
| Result  |
+---------+--+
+---------+--+
No rows selected (2.515 seconds)
0: jdbc:hive2://ha-cluster/default> show tables;
+-----------+------------+--------------+--+
| database  | tableName  | isTemporary  |
+-----------+------------+--------------+--+
| default   | test       | false        |
| default   | test_obs   | false        |
+-----------+------------+--------------+--+
2 rows selected (0.127 seconds)
```

d. Press **Ctrl+C** to exit the Spark beeline.

6. Verify that Presto can access OBS.

   – For normal clusters with Kerberos authentication disabled

     i. Run the following command to connect to the client:

        presto_cli.sh

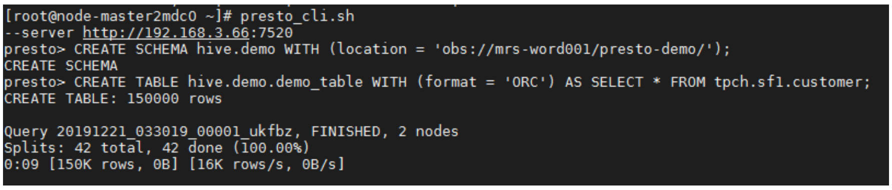
     ii. On the Presto client, run the following statement to create a schema and set **location** to an OBS path:

        CREATE SCHEMA hive.demo01 WITH (location = 'obs://mrs-word001/presto-demo002/');

     iii. Create a table in the schema. The table data is stored in the OBS file system. The following is an example.

        CREATE TABLE hive.demo.demo_table WITH (format = 'ORC') AS SELECT * FROM tpch.sf1.customer;

Figure 6-4 Return result



```
[root@node-master2mdc0 ~]# presto_cli.sh
--server http://192.168.3.66:7520
presto> CREATE SCHEMA hive.demo WITH (location = 'obs://mrs-word001/presto-demo/');
CREATE SCHEMA
presto> CREATE TABLE hive.demo.demo_table WITH (format = 'ORC') AS SELECT * FROM tpch.sf1.customer;
CREATE TABLE: 150000 rows

Query 20191221_033019_00001_ukfbz, FINISHED, 2 nodes
Splits: 42 total, 42 done (100.00%)
0:09 [150K rows, 0B] [16K rows/s, 0B/s]
```

     iv. Run **exit** to exit the client.

   – For security clusters with Kerberos authentication enabled

     i. Log in to MRS Manager and create a role with the Hive Admin Privilege permissions, for example, **prestorole**. For details about how to create a role, see **Creating a Role**.

     ii. Create a user that belongs to the Presto and Hive groups and bind the role created in **6.i** to the user, for example, **presto001**. For details about how to create a user, see **Creating a User**.

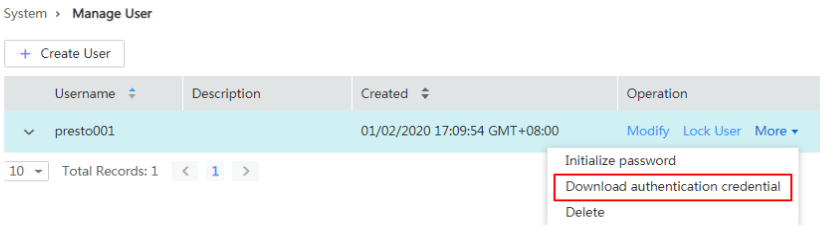
     iii. Authenticate the current user.

        kinit presto001

     iv. Download the user credential.

1) For MRS 3.x earlier, on MRS Manager, choose **System** > **Manage User**. In the row of the new user, choose **More** > **Download Authentication Credential**.

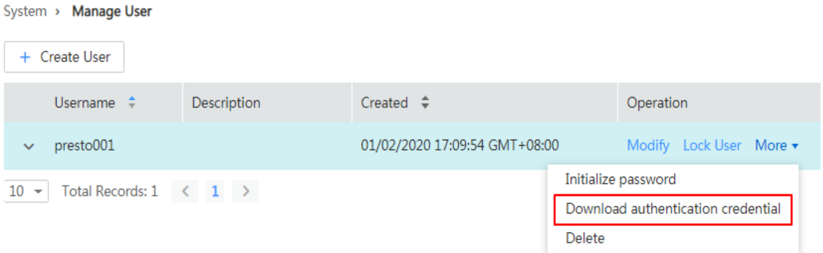
Figure 6-5 Downloading the Presto user authentication credential



Username	Description	Created	Operation
presto001		01/02/2020 17:09:54 GMT+08:00	Modify Lock User More

2) On FusionInsight Manager for MRS 3.x or later,, choose **System > Permission > User**. In the row that contains the newly added user, click **More > Download Authentication Credential**.

Figure 6-6 Downloading the Presto user authentication credential



Username	Description	Created	Operation
presto001		01/02/2020 17:09:54 GMT+08:00	Modify Lock User More

v. Decompress the downloaded user credential file, and save the obtained **krb5.conf** and **user.keytab** files to the client directory, for example, **/opt/Bigdata/client/Presto/**.

vi. Run the following command to obtain a user principal:

klist -kt /opt/Bigdata/client/Presto/user.keytab

vii. For clusters with Kerberos authentication enabled, run the following command to connect to the Presto Server of the cluster:

presto_cli.sh --krb5-config-path {krb5.conf file path} --krb5-principal {user principal} --krb5-keytab-path {user.keytab file path} --user {presto username}

- **krb5.conf** file path: Replace it with the file path set in **6.v**, for example, **/opt/Bigdata/client/Presto/krb5.conf**.
- **user.keytab** file path: Replace it with the file path set in **6.v**, for example, **/opt/Bigdata/client/Presto/user.keytab**.
- **user principal**: Replace it with the result returned in **6.vi**.
- **presto username**: Replace it with the name of the user created in **6.ii**, for example, **presto001**.

Example: presto_cli.sh --krb5-config-path /opt/Bigdata/client/Presto/krb5.conf --krb5-principal prest001@xxx_xxx_xxx_xxx.COM --krb5-keytab-path /opt/Bigdata/client/Presto/user.keytab --user presto001

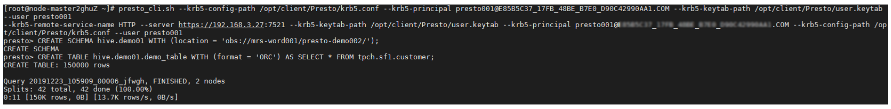
viii. On the Presto client, run the following statement to create a schema and set **location** to an OBS path:

CREATE SCHEMA hive.demo01 WITH (location = 'obs://mrs-word001/presto-demo002/');

ix. Create a table in the schema. The table data is stored in the OBS file system. The following is an example.

CREATE TABLE hive.demo01.demo_table WITH (format = 'ORC') AS SELECT * FROM tpch.sf1.customer;

Figure 6-7 Return result



```
[root@node-master2ghuZ ~]# presto_cli.sh --krb5-config-path /opt/client/Presto/krb5.conf --krb5-principal presto001@E85B5C37_17FB_48BE_B7E0_D90C42990AA1.COM --krb5-keytab-path /opt/client/Presto/user.keytab
--user presto001
--krb5-remote-service-name HTTP --server https://192.168.3.27:7521 --krb5-keytab-path /opt/client/Presto/user.keytab --krb5-principal presto001@[redacted].COM --krb5-config-path /op
t/client/Presto/krb5.conf --user presto001
presto> CREATE SCHEMA hive.demo01 WITH (location = 'obs://mrs-word001/presto-demo002/');
CREATE SCHEMA
presto> CREATE TABLE hive.demo01.demo_table WITH (format = 'ORC') AS SELECT * FROM tpch.sf1.customer;
CREATE TABLE: 150000 rows

Query 20191223_105909_00006_jfwgh, FINISHED, 2 nodes
Splits: 42 total, 42 done (100.00%)
0:11 [150K rows, 0B] [13.7K rows/s, 0B/s]
```

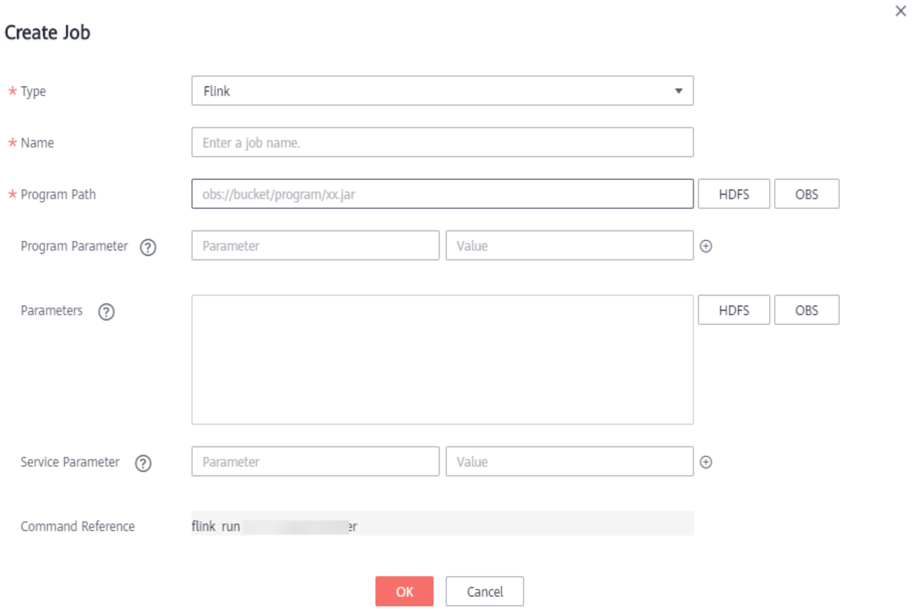
x. Run **exit** to exit the client.

7. Verify that Flink can access OBS.

   a. On the **Dashboard** page, click **Synchronize** on the right of **IAM User Sync** to synchronize IAM users.

   b. After user synchronization is complete, choose **Jobs** > **Create** on the cluster details page to create a Flink job. In **Parameters**, enter parameters in **--input <Job input path> --output <Job output path>** format. You can click **OBS** to select a job input path, and enter a job output path that does not exist, for example, **obs://mrs-word001/output/**. See **Figure 6-8**.

   Figure 6-8 Creating a Flink job



Create Job

* Type: Flink

* Name: Enter a job name.

* Program Path: obs://bucket/program/xx.jar

Program Parameter: Parameter | Value

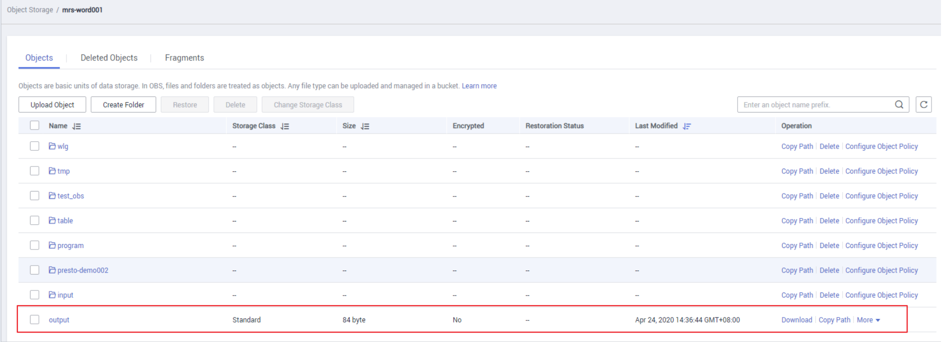
Parameters:

Service Parameter: Parameter | Value

Command Reference: flink run [redacted]er

   c. On OBS Console, go to the output path specified during job creation. If the output directory is automatically created and contains the job execution results, OBS access is successful.

Figure 6-9 Flink job execution result



Object Storage / **mrs-word001**

Objects | Deleted Objects | Fragments

Objects are basic units of data storage. In OBS, files and folders are treated as objects. Any file type can be uploaded and managed in a bucket. Learn more

Name	Storage Class	Size	Encrypted	Restoration Status	Last Modified	Operation
wlg	--	--	--	--	--	Copy Path | Delete | Configure Object Policy
tmp	--	--	--	--	--	Copy Path | Delete | Configure Object Policy
test_obs	--	--	--	--	--	Copy Path | Delete | Configure Object Policy
table	--	--	--	--	--	Copy Path | Delete | Configure Object Policy
program	--	--	--	--	--	Copy Path | Delete | Configure Object Policy
presto-demo002	--	--	--	--	--	Copy Path | Delete | Configure Object Policy
input	--	--	--	--	--	Copy Path | Delete | Configure Object Policy
output	Standard	84 byte	No	--	Apr 24, 2020 14:36:44 GMT+08:00	Download | Copy Path | More

Reference

For details about how to control permissions to access OBS, see **Configuring Fine-Grained Permissions for MRS Multi-User Access to OBS**.

6.3 Configuring a Storage-Compute Decoupled Cluster (AK/SK)

In MRS 2.0.1 or later, OBS can be interconnected with MRS using **obs://**. Currently, Hadoop, Hive, Spark, Presto, and Flink are supported. HBase cannot use **obs://** to interconnect with OBS.

MRS provides the following configuration modes for accessing OBS. You can select one of them. The agency mode is recommended.

- Bind an agency of the ECS type to an MRS cluster to access OBS, preventing the AK/SK from being exposed in the configuration file. For details, see **Configuring a Storage-Compute Decoupled Cluster (Agency)**.
- Configure the AK/SK in an MRS cluster. The AK/SK will be exposed in the configuration file in plaintext. Exercise caution when performing this operation. For details, see the following part in this section.

NOTE

- To improve data write performance, change the value of the **fs.obs.buffer.dir** parameter of the corresponding service to a data disk directory.
- In the big data decoupled storage-compute scenario, the OBS parallel file system must be used to configure a cluster. Using common object buckets will greatly affect the cluster performance.

Using Hadoop to Access OBS

- Add the following content to file **core-site.xml** in the HDFS directory (**$client_home/ HDFS/hadoop/etc/hadoop**) on the MRS client:

```
<property>
   <name>fs.obs.access.key</name>
   <value>ak</value>
</property>
<property>
   <name>fs.obs.secret.key</name>
   <value>sk</value>
</property>
<property>
```

```
    <name>fs.obs.endpoint</name>
    <value>obs endpoint</value>
</property>
```

NOTICE

AK and SK will be displayed as plaintext in the configuration file. Exercise caution when setting AK and SK in the file.

After the configuration is added, you can directly access data on OBS without manually adding the AK/SK and endpoint. For example, run the following command to view the file list of the **test_obs_orc** directory in the **obs-test** file system:

hadoop fs -ls "obs://obs-test/test_obs_orc"

- Add AK/SK and endpoint to the command line to access data on OBS.

  hadoop fs -Dfs.obs.endpoint=xxx -Dfs.obs.access.key=xx -Dfs.obs.secret.key=xx -ls "obs://obs-test/ test_obs_orc"

Using Hive to Access OBS

Step 1 The Hive service configuration page is displayed.

- For versions earlier than MRS 2.0.1, log in to MRS Manager, choose **Services** > **Hive** > **Service Configuration**, and select **All** from the **Basic** drop-down list.
- For MRS 2.0.1 or later, click the cluster name on the MRS console, choose **Components** > **Hive** > **Service Configuration**, and select **All** from the **Basic** drop-down list.

  NOTE

  If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)
- For MRS 3.*x* or later, log in to FusionInsight Manager. For details, see **Accessing FusionInsight Manager (MRS 3.x or Later)**. And choose **Cluster** > *Name of the desired cluster* > **Services** > **Hive** > **Configurations** > **All Configurations**.

Step 2 In the configuration type drop-down box, switch **Basic Configurations** to **All Configurations**.

Step 3 Search for **fs.obs.access.key** and **fs.obs.secret.key** and set them to the AK and SK of OBS respectively.

If the preceding two parameters cannot be found in the current cluster, choose **Hive > Customization** in the navigation tree on the left and add the two parameters to the customized parameter **core.site.customized.configs**.

Step 4 Click **Save Configuration** and select **Restart the affected services or instances**. to restart the Hive service.

Step 5 Access the OBS directory in the beeline. For example, run the following command to create a Hive table and specify that data is stored in the **test_obs** directory in the **test-bucket** file system:

create table test_obs(a int, b string) row format delimited fields terminated by "," stored as textfile location "obs://test-bucket/test_obs";

----End

Using Spark to Access OBS

NOTE

SparkSQL depends on Hive. Therefore, when configuring OBS on Spark, you need to modify the OBS configuration used in **Using Hive to Access OBS**.

- spark-beeline and spark-sql

  You can add the following OBS attributes to the shell to access OBS:

```
set fs.obs.endpoint=xxx
set fs.obs.access.key=xxx
set fs.obs.secret.key=xxx
```

- spark-beeline

  The spark-beeline can access OBS by configuring service parameters on Manager. The procedure is as follows:

  a. Go to the Spark configuration page.

     - For versions earlier than MRS 2.0.1, log in to MRS Manager and choose **Services** > **Spark** > **Service Configuration**.
     - For MRS 2.0.1 or later, click the cluster name on the MRS console and choose **Components** > **Spark** > **Service Configuration**.

       NOTE

       If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

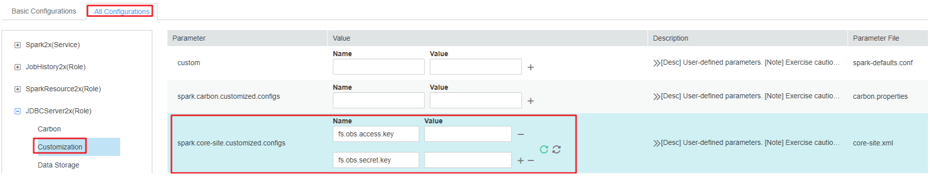
     - For MRS 3.*x* or later, log in to FusionInsight Manager. For details, see **Accessing FusionInsight Manager (MRS 3.x or Later)**. Choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Configurations**.

  b. In the configuration type drop-down box, switch **Basic Configurations** to **All Configurations**.

  c. Choose **JDBCServer** > **OBS**, and set values for **fs.obs.access.key** and **fs.obs.secret.key**.

     If the preceding two parameters cannot be found in the current cluster, choose **JDBCServer** > **Customization** in the navigation tree on the left and add the two parameters to the customized parameter **spark.core-site.customized.configs**.

Figure 6-10 Parameters for adding an OBS



Parameter	Value	Description	Parameter File
custom	Name: Value:	»[Desc] User-defined parameters. [Note] Exercise cautio...	spark-defaults.conf
spark.carbon.customized.configs	Name: Value:	»[Desc] User-defined parameters. [Note] Exercise cautio...	carbon.properties
spark.core-site.customized.configs	Name: fs.obs.access.key Value: Name: fs.obs.secret.key Value:	»[Desc] User-defined parameters. [Note] Exercise cautio...	core-site.xml

   d. Click **Save Configuration** and select **Restart the affected services or instances**. Restart the Spark service.

   e. Access OBS in **spark-beeline**. For example, access the **obs://obs-demo-input/table/** directory.

      create table test(id int) location 'obs://obs-demo-input/table/';

- spark-sql and spark-submit

  The spark-sql can also access OBS by modifying the **core-site.xml** configuration file.

  The method of modifying the configuration file is the same when you use the spark-sql and spark-submit to submit a task to access OBS.

  Add the following content to **core-site.xml** in the Spark configuration folder (**$client_home/Spark/spark/conf**) on the MRS client:

```
<property>
   <name>fs.obs.access.key</name>
   <value>ak</value>
</property>
<property>
   <name>fs.obs.secret.key</name>
   <value>sk</value>
</property>
<property>
   <name>fs.obs.endpoint</name>
   <value>obs endpoint</value>
</property>
```

Using Presto to Access OBS

Step 1 Go to the cluster details page and choose **Components** > **Presto** > **Service Configuration**.

Step 2 In the configuration type drop-down box, switch **Basic Configurations** to **All Configurations**.

Step 3 Search for and configure the following parameters:

- Set **fs.obs.access.key** to **AK**.
- Set **fs.obs.secret.key** to **SK**.

If the preceding two parameters cannot be found in the current cluster, choose **Presto > Hive** in the navigation tree on the left and add the two parameters to the customized parameter **core.site.customized.configs**.

Step 4 Click **Save Configuration** and select **Restart the affected services or instances**. to restart the Presto service.

Step 5 Choose **Components** > **Hive** > **Service Configuration**.

Step 6 In the configuration type drop-down box, switch **Basic Configurations** to **All Configurations**.

Step 7 Search for and configure the following parameters:

- Set **fs.obs.access.key** to **AK**.
- Set **fs.obs.secret.key** to **SK**.

Step 8 Click **Save Configuration** and select **Restart the affected services or instances**. to restart the Hive service.

Step 9 On the Presto client, run the following statement to create a schema and set **location** to an OBS path:

CREATE SCHEMA hive.demo WITH (location = 'obs://obs-demo/presto-demo/');

Step 10 Create a table in the schema. The table data is stored in the OBS file system. The following is an example.

CREATE TABLE hive.demo.demo_table WITH (format = 'ORC') AS SELECT * FROM tpch.sf1.customer;

----End

Using Flink to Access OBS

Add the following configuration to the Flink configuration file of the MRS client in ***Client installation path*/Flink/flink/conf/flink-conf.yaml**:

```
fs.obs.access.key:ak
fs.obs.secret.key: sk
fs.obs.endpoint: obs endpoint
```

NOTICE

AK and SK will be displayed as plaintext in the configuration file. Exercise caution when setting AK and SK in the file.

After the configuration is added, you can directly access data on OBS without manually adding the AK/SK and endpoint.

6.4 Using a Storage-Compute Decoupled Cluster

6.4.1 Interconnecting Flink with OBS

Before performing the following operations, ensure that you have configured a storage-compute decoupled cluster by referring to **Configuring a Storage-Compute Decoupled Cluster (Agency)** or **Configuring a Storage-Compute Decoupled Cluster (AK/SK)**.

Step 1 Log in to the Flink client installation node as the client installation user.

Step 2 Run the following command to initialize environment variables:

source ${client_home}/bigdata_env

Step 3 Configure the Flink client properly. For details, see **Installing a Client (Version 3.x or Later)**.

Step 4 For a security cluster, run the following command to perform user authentication. If Kerberos authentication is not enabled for the current cluster, you do not need to run this command.

kinit *Username*

Step 5 Explicitly add the OBS file system to be accessed in the Flink command line.

./bin/flink run --class com.xxx.bigdata.flink.examples.FlinkProcessingTimeAPIMain ./config/FlinkCheckpointJavaExample.jar --chkPath obs://*Name of the OBS parallel file system*

----End

NOTE

Flink jobs are running on Yarn. Before configuring Flink to interconnect with the OBS file system, ensure that the interconnection between Yarn and the OBS file system is normal.

6.4.2 Interconnecting Flume with OBS

This section applies to MRS 3.x or later.

Before performing the following operations, ensure that you have configured a storage-compute decoupled cluster by referring to **Configuring a Storage-Compute Decoupled Cluster (Agency)** or **Configuring a Storage-Compute Decoupled Cluster (AK/SK)**.

Step 1 Configure an agency.

1. Log in to the MRS console. In the navigation pane on the left, choose **Clusters** > **Active Clusters**.
2. Click the name of a cluster to go to the cluster details page.
3. On the **Dashboard** page, click **Synchronize** on the right of **IAM User Sync** to synchronize IAM users.
4. Click **Manage Agency** on the right of **Agency**, select the target agency, and click **OK**.

Step 2 Create an OBS file system for storing data.

1. Log in to the OBS console.
2. In the navigation pane on the left, choose **Parallel File Systems**. On the displayed page, click **Create Parallel File System**.
3. Enter the file system name, for example, **esdk-c-test-pfs1**, and set other parameters as required. Click **Create Now**.
4. In the parallel file system list on the OBS console, click the created file system name to go to its details page.
5. In the navigation pane on the left, choose **Files** and click **Create Folder** to create the **testFlumeOutput** folder.

Step 3 Prepare the **properties.properties** file and upload it to the **/opt/flumeInput** directory.

1. Prepare the **properties.properties** file on the local host. Its content is as follows:

```
# source
server.sources = r1
# channels
server.channels = c1
# sink
server.sinks = obs_sink
# ----- define net source -----
```

```
server.sources.r1.type = seq
server.sources.r1.spooldir = /opt/flumeInput
# ---- define OBS sink ----
server.sinks.obs_sink.type = hdfs
server.sinks.obs_sink.hdfs.path = obs://esdk-c-test-pfs1/testFlumeOutput
server.sinks.obs_sink.hdfs.filePrefix = %[localhost]
server.sinks.obs_sink.hdfs.useLocalTimeStamp = true
# set file size to trigger roll
server.sinks.obs_sink.hdfs.rollSize = 0
server.sinks.obs_sink.hdfs.rollCount = 0
server.sinks.obs_sink.hdfs.rollInterval = 5
#server.sinks.obs_sink.hdfs.threadsPoolSize = 30
server.sinks.obs_sink.hdfs.fileType = DataStream
server.sinks.obs_sink.hdfs.writeFormat = Text
server.sinks.obs_sink.hdfs.fileCloseByEndEvent = false

# define channel
server.channels.c1.type = memory
server.channels.c1.capacity = 1000
# transaction size
server.channels.c1.transactionCapacity = 1000
server.channels.c1.byteCapacity = 800000
server.channels.c1.byteCapacityBufferPercentage = 20
server.channels.c1.keep-alive = 60
server.sources.r1.channels = c1
server.sinks.obs_sink.channel = c1
```

NOTE

The value of **server.sinks.obs_sink.hdfs.path** is the OBS file system created in **Step 2**.

2. Log in to the node where the Flume client is installed as user **root**.
3. Create the **/opt/flumeInput** directory and create a customized **.txt** file in this directory.
4. Log in to FusionInsight Manager.
5. Choose **Cluster** > *Name of the target cluster* > **Services** > **Flume**. On the displayed page, click **Configurations** and then **Upload File** in the **Value** column corresponding to the **flume.config.file** parameter, upload the **properties.properties** file prepared in **Step 3.1**, and click **Save**.

Step 4 View the result in the OBS system.

1. Log in to the OBS console.
2. Click **Parallel File Systems** and go to the folder created in **Step 2** to view the result.

----End

6.4.3 Interconnecting HDFS with OBS

Before performing the following operations, ensure that you have configured a storage-compute decoupled cluster by referring to **Configuring a Storage-Compute Decoupled Cluster (Agency)** or **Configuring a Storage-Compute Decoupled Cluster (AK/SK)**.

Step 1 Log in to the node on which the HDFS client is installed as a client installation user.

Step 2 Run the following command to switch to the client installation directory.

cd ${client_home}

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 If the cluster is in security mode, run the following command to authenticate the user. In normal mode, skip user authentication.

kinit *Component service user*

Step 5 Explicitly add the OBS file system to be accessed in the HDFS command line.

Example:

- Run the following command to access the OBS file system:

  hdfs dfs -ls obs://*OBS parallel file system name/Path*
- Run the following command to upload the **/opt/test.txt** file from the client node to the OBS file system path:

  hdfs dfs -put /opt/test.txt obs://*OBS parallel file system name/Path*

----End

NOTE

If a large number of logs are printed in the OBS file system, the read and write performance may be affected. You can adjust the log level of the OBS client as follows:

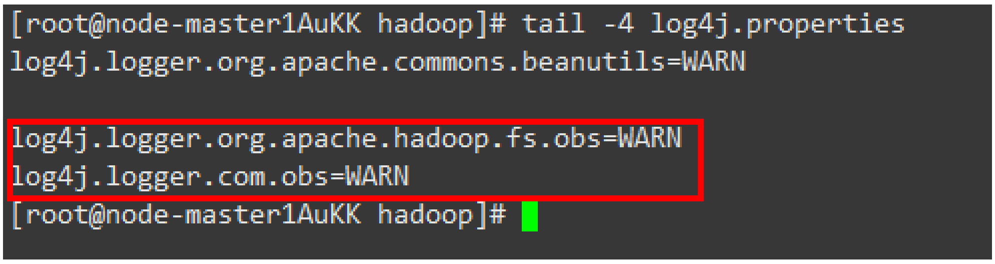
cd ${client_home}/HDFS/hadoop/etc/hadoop

vi log4j.properties

Add the OBS log level configuration to the file as follows:

log4j.logger.org.apache.hadoop.fs.obs=WARN

log4j.logger.com.obs=WARN



```
[root@node-master1AuKK hadoop]# tail -4 log4j.properties
log4j.logger.org.apache.commons.beanutils=WARN

log4j.logger.org.apache.hadoop.fs.obs=WARN
log4j.logger.com.obs=WARN
[root@node-master1AuKK hadoop]#
```

6.4.4 Interconnecting Hive with OBS

Before performing the following operations, ensure that you have configured a storage-compute decoupled cluster by referring to **Configuring a Storage-Compute Decoupled Cluster (Agency)** or **Configuring a Storage-Compute Decoupled Cluster (AK/SK)**.

When creating a table, set the table location to an OBS path.

Step 1 Log in to the client installation node as the client installation user.

Step 2 Run the following command to initialize environment variables:

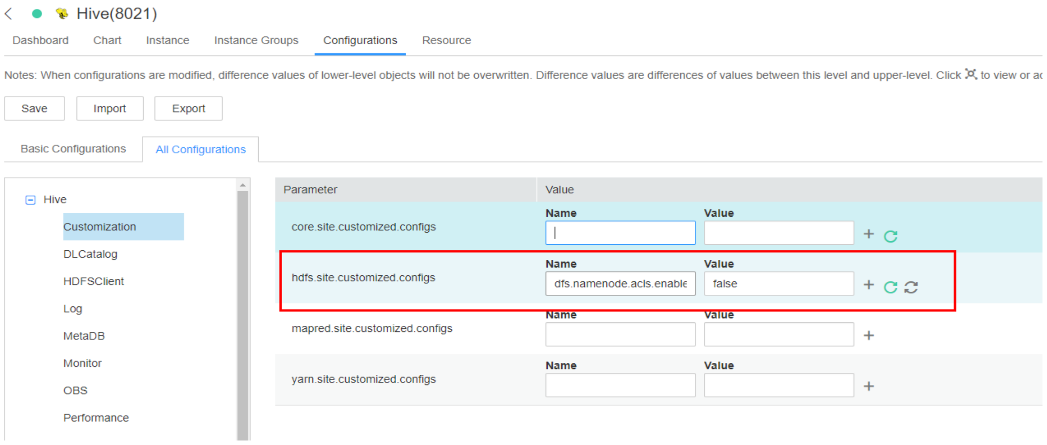
source ${client_home}/bigdata_env

Step 3 For a security cluster, run the following command to perform user authentication (the user must have the permission to perform Hive operations). If Kerberos authentication is not enabled for the current cluster, you do not need to run this command.

kinit *User performing Hive operations*

Step 4 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **Hive** > **Configurations** > **All Configurations**.

In the left navigation tree, choose **Hive** > **Customization**. In the customized configuration items, add **dfs.namenode.acls.enabled** to the **hdfs.site.customized.configs** parameter and set its value to **false**.



Notes: When configurations are modified, difference values of lower-level objects will not be overwritten. Difference values are differences of values between this level and upper-level. Click to view or ac

Parameter	Value: Name	Value: Value
core.site.customized.configs		
hdfs.site.customized.configs	dfs.namenode.acls.enable	false
mapred.site.customized.configs		
yarn.site.customized.configs		

Step 5 Click **Save**. Click the **Dashboard** tab and choose **More** > **Restart Service**. In the **Verify Identity** dialog box that is displayed, enter the password of the current user, and click **OK**. In the displayed **Restart Service** dialog box, select **Restart upper-layer services** and click **OK**. Hive is restarted.

Step 6 Log in to the beeline client and set **Location** to the OBS file system path when creating a table.

beeline

For example, run the following command to create the table **test** in **obs://***OBS parallel file system name***/user/hive/warehouse/***Database name***/***Table name*.

create table test(name string) location "obs://*OBS parallel file system name***/user/hive/warehouse/***Database name***/***Table name***";**

NOTE

You need to add the component operator to the URL policy in the Ranger policy. Set the URL to the complete path of the object on OBS. Select the Read and Write permissions.

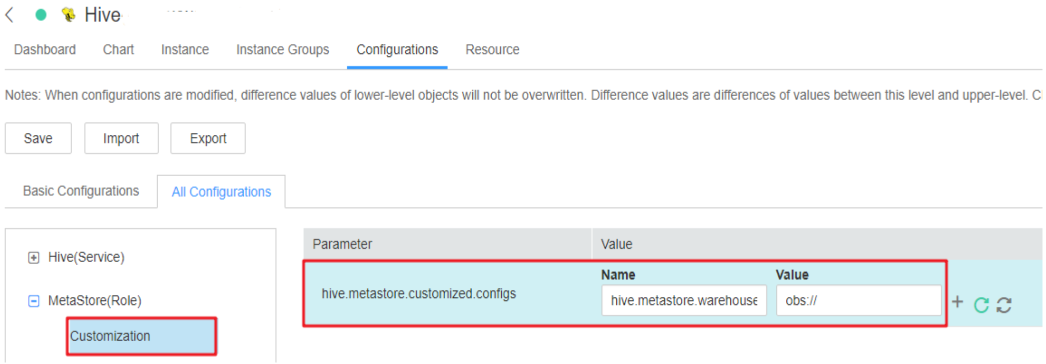
----End

Setting the Default Location of the Created Hive Table to the OBS Path

Step 1 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **Hive** > **Configurations** > **All Configurations**.

Step 2 In the left navigation tree, choose **MetaStore** > **Customization**. Add **hive.metastore.warehouse.dir** to the **hive.metastore.customized.configs** parameter and set it to the OBS path.

Figure 6-11 Configurations of **hive.metastore.warehouse.dir**

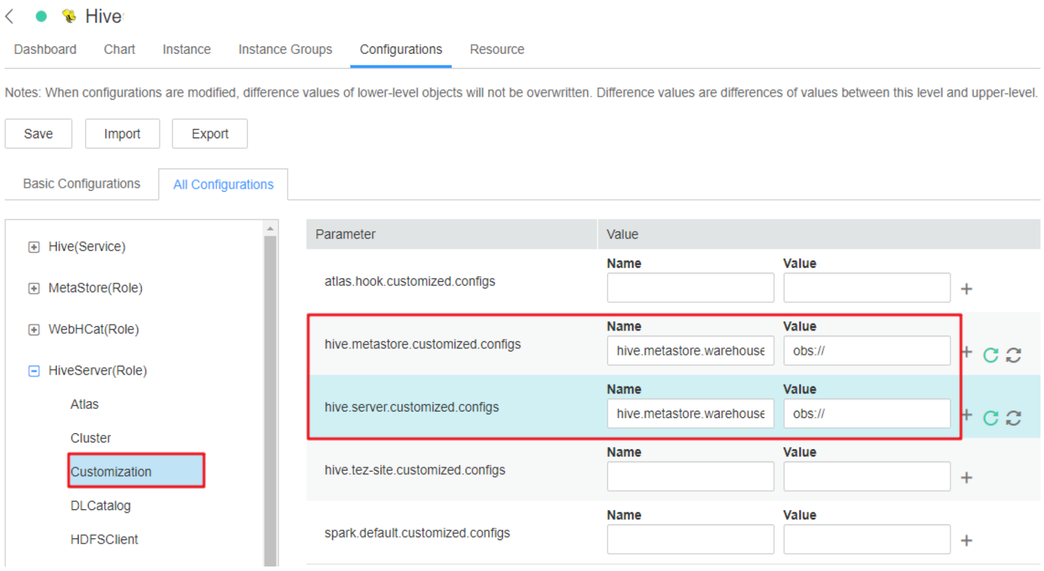


Notes: When configurations are modified, difference values of lower-level objects will not be overwritten. Difference values are differences of values between this level and upper-level. C

Parameter	Name	Value
hive.metastore.customized.configs	hive.metastore.warehouse	obs://

Step 3 In the left navigation tree, choose **HiveServer** > **Customization**. Add **hive.metastore.warehouse.dir** to the **hive.metastore.customized.configs** and **hive.metastore.customized.configs** parameters, and set it to the OBS path.

Figure 6-12 hive.metastore.warehouse.dir configuration



Notes: When configurations are modified, difference values of lower-level objects will not be overwritten. Difference values are differences of values between this level and upper-level.

Parameter	Name	Value
atlas.hook.customized.configs		
hive.metastore.customized.configs	hive.metastore.warehouse	obs://
hive.server.customized.configs	hive.metastore.warehouse	obs://
hive.tez-site.customized.configs		
spark.default.customized.configs		

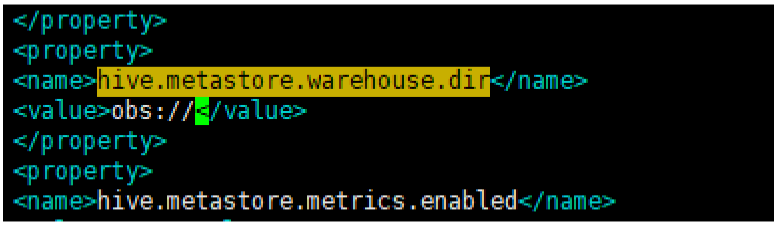
Step 4 Save the configurations and restart Hive.

Step 5 Update the client configuration file.

1. Run the following command to open **hivemetastore-site.xml** in the Hive configuration file directory on the client:

   vim /opt/Bigdata/client/Hive/config/hivemetastore-site.xml

2. Change the value of **hive.metastore.warehouse.dir** to the corresponding OBS path.



```
</property>
<property>
<name>hive.metastore.warehouse.dir</name>
<value>obs://</value>
</property>
<property>
<name>hive.metastore.metrics.enabled</name>
```

Step 6 Log in to the beeline client, create a table, and check whether the location is the OBS path.

beeline

create table test(name string);

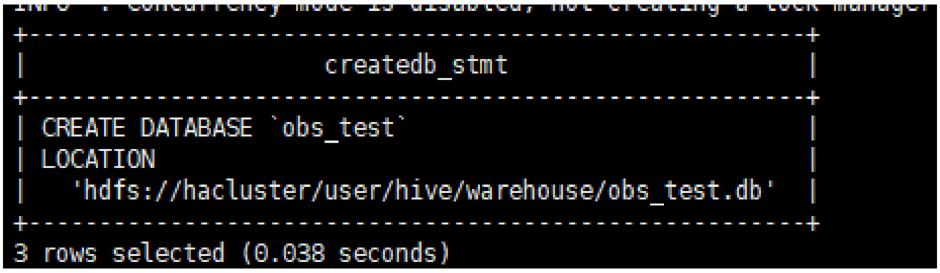
desc formatted test;

NOTE

If the database location points to HDFS, the table to be created in the database (without specifying the location) also points to HDFS. If you want to modify the default table creation policy, change the location of the database to OBS by performing the following operations:

1. Run the following command to query the location of the database:

   show create database *obs_test***;**

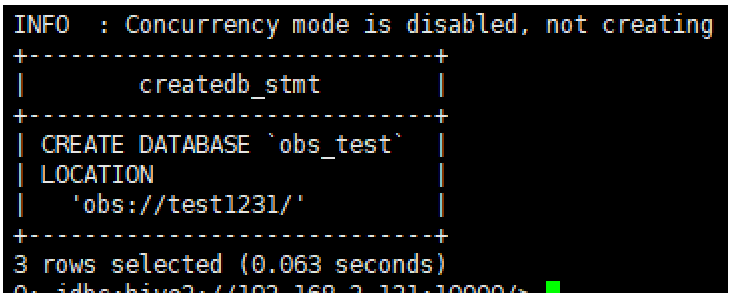


```
INFO  : Concurrency mode is disabled, not creating a lock manager
+----------------------------------------------------+
|                    createdb_stmt                   |
+----------------------------------------------------+
| CREATE DATABASE `obs_test`                         |
| LOCATION                                           |
|   'hdfs://hacluster/user/hive/warehouse/obs_test.db'  |
+----------------------------------------------------+
3 rows selected (0.038 seconds)
```

2. Run the following command to change the database location:

   alter database *obs_test* **set location** '**obs://***OBS parallel file system name***/user/hive/warehouse/***Database name*'

   Run the **show create database** *obs_test* command to check whether the database location points to OBS.



```
INFO  : Concurrency mode is disabled, not creating
+---------------------------+
|        createdb_stmt      |
+---------------------------+
| CREATE DATABASE `obs_test`  |
| LOCATION                  |
|   'obs://test1231/'       |
+---------------------------+
3 rows selected (0.063 seconds)
0: jdbc:hive2://192.168.2.121:10000/>
```

3. Run the following command to change the table location:

   alter table *user_info* **set location** '**obs://***OBS parallel file system name***/user/hive/warehouse/***Database name***/***Table name*'

   If the table contains data, migrate the original data file to the new location.

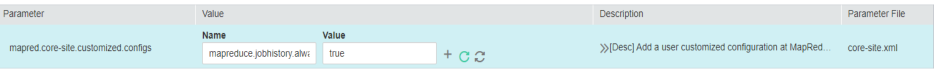
----End

6.4.5 Interconnecting MapReduce with OBS

Before performing the following operations, ensure that you have configured a storage-compute decoupled cluster by referring to **Configuring a Storage-Compute Decoupled Cluster (Agency)** or **Configuring a Storage-Compute Decoupled Cluster (AK/SK)**.

Step 1 Log in to the MRS management console and click the cluster name to go to the cluster details page.

Step 2 Choose **Components > MapReduce**. The **All Configurations** page is displayed. In the navigation tree on the left, choose **MapReduce > Customization**. In the customized configuration items, add the configuration item **mapreduce.jobhistory.always-scan-user-dir** to **core-site.xml** and set its value to **true**.



Parameter	Value	Description	Parameter File
mapred.core-site.customized.configs	Name: mapreduce.jobhistory.alwa Value: true	»[Desc] Add a user customized configuration at MapRed...	core-site.xml

Step 3 Save the configurations and restart the MapReduce service.

----End

6.4.6 Interconnecting Spark2x with OBS

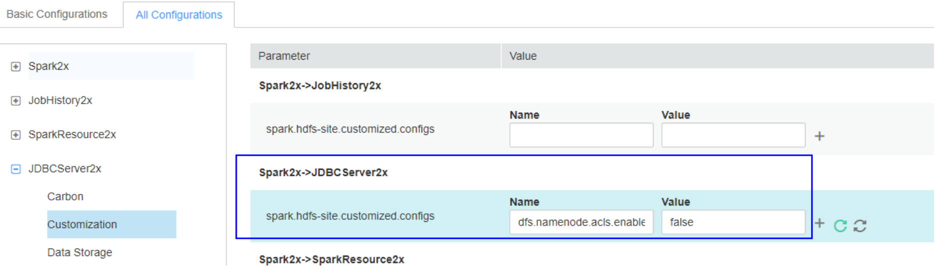
The OBS file system can be interconnected with Spark2x after an MRS cluster is installed.

Before performing the following operations, ensure that you have configured a storage-compute decoupled cluster by referring to **Configuring a Storage-Compute Decoupled Cluster (Agency)** or **Configuring a Storage-Compute Decoupled Cluster (AK/SK)**.

Using Spark Beeline After Cluster Installation

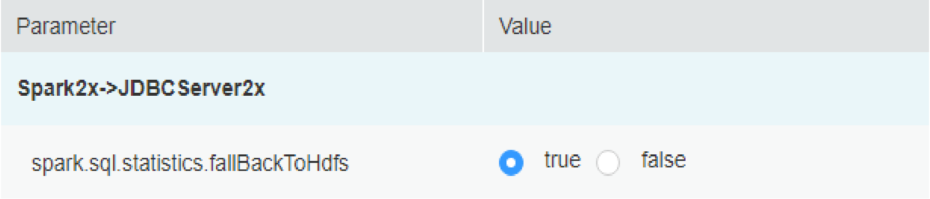
Step 1 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **Spark2x** > **Configurations** > **All Configurations**.

In the left navigation tree, choose **JDBCServer2x** > **Customization**. Add **dfs.namenode.acls.enabled** to the **spark.hdfs-site.customized.configs** parameter and set its value to **false**.



Parameter	Value
Spark2x->JobHistory2x	
spark.hdfs-site.customized.configs	Name: Value:
Spark2x->JDBCServer2x	
spark.hdfs-site.customized.configs	Name: dfs.namenode.acls.enable Value: false
Spark2x->SparkResource2x	

Step 2 Search for the **spark.sql.statistics.fallBackToHdfs** parameter and set its value to **false**.



Parameter	Value
Spark2x->JDBCServer2x	
spark.sql.statistics.fallBackToHdfs	● true ○ false

Step 3 Save the configurations and restart the JDBCServer2x instance.

Step 4 Log in to the client installation node as the client installation user.

Step 5 Run the following command to configure environment variables:

source ${client_home}/bigdata_env

Step 6 For a security cluster, run the following command to perform user authentication. If Kerberos authentication is not enabled for the current cluster, you do not need to run this command.

kinit *Username*

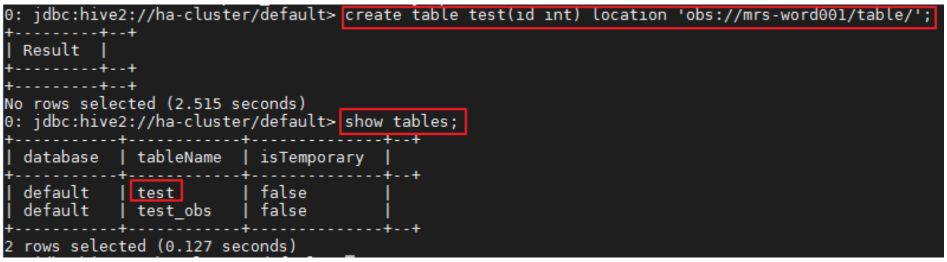
Step 7 Access OBS in spark-beeline. For example, create a table named **test** in the **obs://mrs-word001/table/** directory.

create table test(id int) location '*obs://mrs-word001/table/***';**

Step 8 Run the following command to query all tables. If table **test** is displayed in the command output, OBS access is successful.

show tables;

Figure 6-13 Verifying the created table name returned using Spark2x



```
0: jdbc:hive2://ha-cluster/default> create table test(id int) location 'obs://mrs-word001/table/';
+---------+--+
| Result  |
+---------+--+
+---------+--+
No rows selected (2.515 seconds)
0: jdbc:hive2://ha-cluster/default> show tables;
+-----------+-----------+--------------+--+
| database  | tableName | isTemporary  |
+-----------+-----------+--------------+--+
| default   | test      | false        |
| default   | test_obs  | false        |
+-----------+-----------+--------------+--+
2 rows selected (0.127 seconds)
```

Step 9 Press **Ctrl+C** to exit the Spark Beeline.

----End

Using Spark SQL After Cluster Installation

Step 1 Log in to the client installation node as the client installation user.

Step 2 Run the following command to configure environment variables:

source ${client_home}/bigdata_env

Step 3 Modify the configuration file:

vim ${client_home}/Spark2x/spark/conf/hdfs-site.xml

```
<property>
<name>dfs.namenode.acls.enabled</name>
<value>false</value>
</property>
```

Step 4 For a security cluster, run the following command to perform user authentication. If Kerberos authentication is not enabled for the current cluster, you do not need to run this command.

kinit *Username*

Step 5 Access OBS in spark-sql. For example, create a table named **test** in the **obs://mrs-word001/table/** directory.

Step 6 Run the **cd ${client_home}/Spark2x/spark/bin** command to access the **spark bin** directory and run **./spark-sql** to log in to spark-sql CLI.

Step 7 Run the following command in the spark-sql CLI:

create table test(id int) location '*obs://mrs-word001/table/***';**

Step 8 Run the **show tables;** command to confirm that the table is created successfully.

Step 9 Run **exit;** to exit the spark-sql CLI.

NOTE

If a large number of logs are printed in the OBS file system, the read and write performance may be affected. You can adjust the log level of the OBS client as follows:

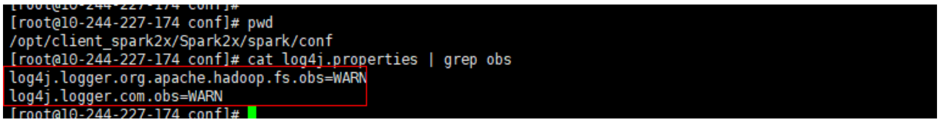
cd ${client_home}/Spark2x/spark/conf

vi log4j.properties

Add the OBS log level configuration to the file as follows:

log4j.logger.org.apache.hadoop.fs.obs=WARN

log4j.logger.com.obs=WARN



```
[root@10-244-227-174 conf]#
[root@10-244-227-174 conf]# pwd
/opt/client_spark2x/Spark2x/spark/conf
[root@10-244-227-174 conf]# cat log4j.properties | grep obs
log4j.logger.org.apache.hadoop.fs.obs=WARN
log4j.logger.com.obs=WARN
[root@10-244-227-174 conf]#
```

----End

6.4.7 Interconnecting Sqoop with External Storage Systems

Exporting Data From HDFS to MySQL Using the sqoop export Command

Step 1 Log in to the node where the client is located.

Step 2 Run the following command to initialize environment variables:

source /opt/client/bigdata_env

Step 3 Run the following command to operate the Sqoop client:

sqoop export --connect jdbc:mysql://10.100.231.134:3306/test --username root --password xxxxxx --table component13 -export-dir hdfs://hacluster/user/hive/warehouse/component_test3 --fields-terminated-by ',' -m 1

Table 6-1 Parameter description

Parameter	Description
-direct	Imports data to a relational database using a database import tool, for example, mysqlimport of MySQL, more efficient than the JDBC connection mode.
-export-dir <dir>	Specifies the source directory for storing data in the HDFS.
-m or -num-mappers <n>	Starts *n* (4 by default) maps to import data concurrently. The value cannot be greater than the maximum number of maps in a cluster.
-table <table-name>	Specifies the relational database table to be imported.

Parameter	Description
-update-key <col-name>	Specifies the column used for updating the existing data in a relational database.
-update-mode <mode>	Specifies how updates are performed. The value can be **updateonly** or **allowinsert**. This parameter is used only when the relational data table does not contain the data record to be imported. For example, if the HDFS data to be imported to the destination table contains a data record **id=1** and the table contains an existing data record **id=2**, the update will fail.
-input-null-string <null-string>	This parameter is optional. If it is not specified, **null** will be used.
-input-null-non-string <null-string>	This parameter is optional. If it is not specified, **null** will be used.
-staging-table <staging-table-name>	Creates a table with the same data structure as the destination table for storing data before it is imported to the destination table. This parameter ensures the transaction security when data is imported to a relational database table. Due to multiple transactions during an import, this parameter can prevent other transactions from being affected when one transaction fails. For example, the imported data is incorrect or duplicate records exist.
-clear-staging-table	Clears data in the staging table before data is imported if the staging-table is not empty.

----End

Importing Data from MySQL to Hive Using the sqoop import Command

Step 1 Log in to the node where the client is located.

Step 2 Run the following command to initialize environment variables:

source /opt/client/bigdata_env

Step 3 Run the following command to operate the Sqoop client:

sqoop import --connect jdbc:mysql://10.100.231.134:3306/test --username root --password xxxxxx --table component --hive-import --hive-table component_test2 --delete-target-dir --fields-terminated-by "," -m 1 --as-textfile

Table 6-2 Parameter description

Parameter	Description
-append	Appends data to an existing dataset in the HDFS. Once this parameter is used, Sqoop imports data to a temporary directory, renames the temporary file where the data is stored, and moves the file to a formal directory to avoid duplicate file names in the directory.
-as-avrodatafile	Imports data to a data file in the Avro format.
-as-sequencefile	Imports data to a sequence file.
-as-textfile	Import data to a text file. After the text file is generated, you can run SQL statements in Hive to query the result.
-boundary-query <statement>	Specifies the SQL statement for performing boundary query. Before importing data, use a SQL statement to obtain a result set and import the data in the result set. The data format can be **-boundary-query 'select id,creationdate from person where id = 3'** (indicating a data record whose ID is 3) or **select min(<split-by>), max(<split-by>) from <table name>**. The fields to be queried cannot contain fields whose data type is string. Otherwise, the error message "java.sql.SQLException: Invalid value for getLong()" is displayed.
-columns<col,col,col...>	Specifies the fields to be imported. The format is ***-Column id,Username***.
-direct	Imports data to a relational database using a database import tool, for example, mysqlimport of MySQL, more efficient than the JDBC connection mode.
-direct-split-size	Splits the imported streams by byte. Especially when data is imported from PostgreSQL using the direct mode, a file that reaches the specified size can be divided into several independent files.
-inline-lob-limit	Sets the maximum value of an inline LOB.
-m or -num-mappers	Starts *n* (4 by default) maps to import data concurrently. The value cannot be greater than the maximum number of maps in a cluster.
-query, -e<statement>	Imports data from the query result. To use this parameter, you must specify the **-target-dir** and **-hive-table** parameters and use the query statement containing the WHERE clause as well as $CONDITIONS. Example: **-query'select * from person where $CONDITIONS' -target-dir /user/hive/warehouse/person -hive-table person**

Parameter	Description
-split-by<column-name>	Specifies the column of a table used to split work units. Generally, the column name is followed by the primary key ID.
-table <table-name>	Specifies the relational database table from which data is obtained.
-target-dir <dir>	Specifies the HDFS path.
-warehouse-dir <dir>	Specifies the directory for storing data to be imported. This parameter is applicable when data is imported to HDFS but cannot be used when you import data to Hive directories. This parameter cannot be used together with **-target-dir**.
-where	Specifies the WHERE clause when data is imported from a relational database, for example, **-where 'id = 2'**.
-z,-compress	Compresses sequence, text, and Avro data files using the GZIP compression algorithm. Data is not compressed by default.
–compression-codec	Specifies the Hadoop compression codec. GZIP is used by default.
–null-string <null-string>	Specifies the string to be interpreted as **NULL** for string columns.
–null-non-string<null-string>	Specifies the string to be interpreted as null for non-string columns. If this parameter is not specified, **NULL** will be used.
-check-column (col)	Specifies the column for checking incremental data import, for example, **id**.
-incremental (mode) append or last modified	Incrementally imports data. **append**: appends records, for example, appending records that are greater than the value specified by **last-value**. **lastmodified**: appends data that is modified after the date specified by **last-value**.
-last-value (value)	Specifies the maximum value (greater than the specified value) of the column after the last import. This parameter can be set as required.

----End

Sqoop Usage Example

- Importing data from MySQL to HDFS using the **sqoop import** command

sqoop import --connect jdbc:mysql://10.100.231.134:3306/test --username root --password *xxx* **--query 'SELECT * FROM component where $CONDITIONS and component_id ="MRS 1.0_002"' --target-dir /tmp/component_test --delete-target-dir --fields-terminated-by "," -m 1 --as-textfile**

- Exporting data from OBS to MySQL using the **sqoop export** command

  sqoop export --connect jdbc:mysql://10.100.231.134:3306/test --username root --password *xxx* **--table component14 -export-dir obs://obs-file-bucket/xx/part-m-00000 --fields-terminated-by ',' -m 1**

- Importing data from MySQL to OBS using the **sqoop import** command

  sqoop import --connect jdbc:mysql://10.100.231.134:3306/test --username root --password *xxx* **--table component --target-dir obs://obs-file-bucket/xx --delete-target-dir --fields-terminated-by "," -m 1 --as-textfile**

- Importing data from MySQL to OBS tables outside Hive

  sqoop import --connect jdbc:mysql://10.100.231.134:3306/test --username root --password *xxx* **--table component --hive-import --hive-table component_test01 --fields-terminated-by "," -m 1 --as-textfile**

6.4.8 Interconnecting Hudi with OBS

Step 1 Log in to the client installation node as the client installation user.

Step 2 Run the following commands to configure environment variables:

source ${client_home}/bigdata_env

source ${client_home}/Hudi/component_env

Step 3 Modify the configuration file:

vim ${client_home}/Hudi/hudi/conf/hdfs-site.xml

```
<property>
<name>dfs.namenode.acls.enabled</name>
<value>false</value>
</property>
```

Step 4 For a security cluster, run the following command to perform user authentication. If Kerberos authentication is not enabled for the current cluster, you do not need to run this command.

kinit *Username*

Step 5 Start spark-shell and run the following commands to create a COW table and save it in OBS:

import org.apache.hudi.QuickstartUtils._

import scala.collection.JavaConversions._

import org.apache.spark.sql.SaveMode._

import org.apache.hudi.DataSourceReadOptions._

import org.apache.hudi.DataSourceWriteOptions._

import org.apache.hudi.config.HoodieWriteConfig._

val tableName = "hudi_cow_table"

```
val basePath = "obs://testhudi/cow_table/"
val dataGen = new DataGenerator
val inserts = convertToStringList(dataGen.generateInserts(10))
val df = spark.read.json(spark.sparkContext.parallelize(inserts, 2))
df.write.format("org.apache.hudi").
options(getQuickstartWriteConfigs).
option(PRECOMBINE_FIELD_OPT_KEY, "ts").
option(RECORDKEY_FIELD_OPT_KEY, "uuid").
option(PARTITIONPATH_FIELD_OPT_KEY, "partitionpath").
option(TABLE_NAME, tableName).
mode(Overwrite).
save(basePath);
```

Step 6 Use DataSource to check whether the table is successfully created and whether the data is normal.

```
val roViewDF = spark.
read.
format("org.apache.hudi").
load(basePath + "/*/*/*/*")
roViewDF.createOrReplaceTempView("hudi_ro_table")
spark.sql("select * from hudi_ro_table").show()
```

Step 7 Run the **:q** command to exit the spark-shell CLI.

----End

7 Accessing Web Pages of Open Source Components Managed in MRS Clusters

7.1 Web UIs of Open Source Components

Scenario

Web UIs of different components are created and hosted on the Master or Core nodes in the MRS cluster by default. You can view information about the components on these web UIs.

Procedure for accessing the web UIs of open-source component:

1. Select an access method.

   MRS provides the following methods for accessing the web UIs of open-source components:

   - **EIP-based Access**: This method is recommended because it is easy to bind an EIP to a cluster.
   - **Creating an SSH Channel for Connecting to an MRS Cluster and Configuring the Browser**: Use this method when the user and the MRS cluster are on different networks.
2. Access the web UIs. For details, see .

Web UIs

NOTE

For clusters with Kerberos authentication enabled, user **admin** does not have the management permission on each component. To access the web UI of each component, create a user who has the management permission on the corresponding component.

Table 7-1 Web UI addresses of open-source components

Cluster Type	Web UI Type	Web UI Address
All Types	MRS Manager	● Applicable to clusters of all versions **https://***Floating IP address of Manager***:28443/web** **NOTE** 1. Ensure that the local host can communicate with the MRS cluster. 2. Log in to the Master2 node remotely, and run the **ifconfig** command. In the command output, **eth0:wsom** indicates the floating IP address of MRS Manager. Record the value of **inet**. If the floating IP address of MRS Manager cannot be queried on the Master2 node, switch to the Master1 node to query and record the floating IP address. If there is only one Master node, query and record the cluster manager IP address of the Master node. ● For versions earlier than MRS 3.x: https://<*EIP*>:9022/mrsmanager?locale=en-us For details, see **Accessing MRS Manager MRS 2.x or Earlier)**. ● For MRS 3.*x* or later, see **Accessing FusionInsight Manager (MRS 3.x or Later)**.
Analysis cluster	HDFS NameNode	● Versions earlier than MRS 3.x: On the cluster details page, choose **Components** > **HDFS** > **NameNode Web UI** > **NameNode (Active)**. ● MRS 3.x or later: On the Manager homepage, choose **Cluster** > **Services** > **HDFS** > **NameNode Web UI** > **NameNode (*Host name*, Active)**.

Cluster Type	Web UI Type	Web UI Address
	HBase HMaster	● Versions earlier than MRS 3.x: On the cluster details page, choose **Components** > **HBase** > **HMaster Web UI** > **HMaster (Active)**. ● MRS 3.x or later: On the Manager homepage, choose **Cluster** > **Services** > **HBase** > **HMaster Web UI** > **HMaster (*Host name*, Active)**.
	MapReduce JobHistoryServer	● Versions earlier than MRS 3.x: On the cluster details page, choose **Components** > **MapReduce** > **JobHistoryServer Web UI** > **JobHistoryServer**. ● MRS 3.x or later: On the Manager homepage, choose **Cluster** > **Services** > **MapReduce** > **JobHistoryServer Web UI** > **JobHistoryServer (*Host name*, Active)**.
	YARN ResourceManager	● Versions earlier than MRS 3.x: On the cluster details page, choose **Components** > **Yarn** > **ResourceManager Web UI** > **ResourceManager (Active)**. ● MRS 3.x or later: On the Manager homepage, choose **Cluster** > **Services** > **Yarn** > **ResourceManager Web UI** > **ResourceManager (*Host name*, Active)**.
	Spark JobHistory	● Versions earlier than MRS 3.x: On the cluster details page, choose **Components** > **Spark** > **Spark Web UI** > **JobHistory**. ● MRS 3.x or later: On the Manager homepage, choose **Cluster** > **Services** > **Spark2x** > **Spark2x Web UI** > **JobHistory2x (*Host name*, Active)**.

Cluster Type	Web UI Type	Web UI Address
	Hue	● Versions earlier than MRS 3.x: On the cluster details page, choose **Components** > **Hue** > **Hue Web UI** > **Hue (Active)**. ● MRS 3.x or later: On the Manager homepage, choose **Cluster** > **Services** > **Hue** > **Hue Web UI** > **Hue (*Host name*, Active)**. Loader is a graphical data migration management tool based on the open-source Sqoop web UI, and its interface is hosted on the Hue web UI.
	Tez	● Versions earlier than MRS 3.x: On the cluster details page, choose **Components** > **Tez** > **Tez Web UI** > **TezUI**. ● MRS 3.x or later: On the Manager homepage, choose **Cluster** > **Services** > **Tez** > **Tez Web UI** > **TezUI (*Host name*, Active)**.
	Presto	● Versions earlier than MRS 3.x: On the cluster details page, choose **Components** > **Presto** > **Presto Web UI** > **Coordinator (Active)**. ● On the Manager homepage, choose **Cluster** > **Services** > **Presto** > **Coordinator Web UI** > **Coordinator (Coordinator)**.
	Ranger	● Versions earlier than MRS 3.x: On the cluster details page, choose **Components** > **Ranger** > **Ranger Web UI** > **RangerAdmin (Active)**. ● MRS 3.x or later: On the Manager homepage, choose **Cluster** > **Services** > **Ranger** > **Ranger Web UI** > **RangerAdmin**.
Stream processing cluster	Storm	● Versions earlier than MRS 3.x: On the cluster details page, choose **Components** > **Storm** > **Storm Web UI** > **UI**. ● On the Manager homepage, choose **Cluster** > **Services** > **Storm** > **Storm Web UI** > **UI (*Host name*)**.

7.2 List of Open Source Component Ports

Common HBase Ports

The protocol type of all ports in the table is TCP.

Parameter	Default Port	Port Description
hbase.master.port	16000	HMaster RPC port. This port is used to connect the HBase client to HMaster. **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
hbase.master.info.port	16010	HMaster HTTPS port. This port is used by the remote web client to connect to the HMaster UI. **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
hbase.regionserver.port	16020	RegoinServer (RS) RPC port. This port is used to connect the HBase client to RegionServer. **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Parameter	Default Port	Port Description
hbase.regionserver.info.port	16030	HTTPS port of the Region server. This port is used by the remote web client to connect to the RegionServer UI. **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
hbase.thrift.info.port	9095	Thrift Server listening port of Thrift Server This port is used for: Listening when the client is connected **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
hbase.regionserver.thrift.port	9090	Thrift Server listening port of RegionServer This port is used for: Listening when the client is connected to the RegionServer **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
hbase.rest.info.port	8085	Port of the RegionServer RESTServer native web page
-	21309	REST port of RegionServer RESTServer

Common HDFS Ports

The protocol type of all ports in the table is TCP.

Parameter	Default Port	Port Description
dfs.namenode.rpc.port	● 9820 (versions earlier than MRS 3.*x*) ● 8020 (MRS 3.*x* and later)	NameNode RPC port This port is used for: 1. Communication between the HDFS client and NameNode 2. Connection between the DataNode and NameNode **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
dfs.namenode.http.port	9870	HDFS HTTP port (NameNode) This port is used for: 1. Point-to-point NameNode checkpoint operations. 2. Connecting the remote web client to the NameNode UI **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
dfs.namenode.https.port	9871	HDFS HTTPS port (NameNode) This port is used for: 1. Point-to-point NameNode checkpoint operations 2. Connecting the remote web client to the NameNode UI **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Parameter	Default Port	Port Description
dfs.datanode.ipc.port	9867	IPC server port of DataNode This port is used for: Connection between the client and DataNode to perform RPC operations. **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
dfs.datanode.port	9866	DataNode data transmission port This port is used for: 1. Transmitting data from HDFS client from or to the DataNode 2. Point-to-point DataNode data transmission **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
dfs.datanode.http.port	9864	DataNode HTTP port This port is used for: Connecting to the DataNode from the remote web client in security mode **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Parameter	Default Port	Port Description
dfs.datanode.https.port	9865	HTTPS port of DataNode This port is used for: Connecting to the DataNode from the remote web client in security mode **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
dfs.JournalNode.rpc.port	8485	RPC port of JournalNode This port is used for: Client communication to access multiple types of information **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
dfs.journalnode.http.port	8480	JournalNode HTTP port This port is used for: Connecting to the JournalNode from the remote web client in security mode **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Parameter	Default Port	Port Description
dfs.journalnode.https.port	8481	HTTPS port of JournalNode This port is used for: Connecting to the JournalNode from the remote web client in security mode **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
httpfs.http.port	14000	Listening port of the HttpFS HTTP server This port is used for: Connecting to the HttpFS from the remote REST API **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Common Hive Ports

The protocol type of all ports in the table is TCP.

Parameter	Default Port	Port Description
templeton.port	9111	Port used for WebHCat to provide the REST service This port is used for: Communication between the WebHCat client and WebHCat server ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Parameter	Default Port	Port Description
hive.server2.thrift.port	10000	Port for HiveServer to provide Thrift services This port is used for: Communication between the HiveServer and HiveServer client ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
hive.metastore.port	9083	Port for MetaStore to provide Thrift services This port is used for: Communication between the MetaStore client and MetaStore, that is, communication between HiveServer and MetaStore. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
hive.server2.webui.port	10002	Web UI port of Hive This port is used for: HTTPS/HTTP communication between Web requests and the Hive UI server

Common Hue Ports

The protocol type of all ports in the table is TCP.

Parameter	Default Port	Port Description
HTTP_PORT	8888	Port for Hue to provide HTTPS services This port is used to provide web services in HTTPS mode, which can be changed. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Common Kafka Ports

The protocol type of all ports in the table is TCP.

Parameter	Default Port	Port Description
port	9092	Port for a broker to receive data and obtain services
ssl.port	9093	SSL port used by a broker to receive data and obtain services
sasl.port	21007	SASL security authentication port provided by a broker, which provides the secure Kafka service
sasl-ssl.port	21009	Port used by a broker to provide encrypted service based on the SASL and SSL protocols

Common Loader Ports

The protocol type of all ports in the table is TCP.

Parameter	Default Port	Port Description
LOADER_HTTPS_PORT	21351	This port is used to provide REST APIs for configuration and running of Loader jobs. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Common Manager Ports

The protocol type of all ports in the table is TCP.

Parameter	Default Port	Port Description
-	8080	Port provided by WebService for user access This port is used to access the web UI over HTTP. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Parameter	Default Port	Port Description
-	28443	Port provided by WebService for user access This port is used to access the web UI over HTTPS. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Common MapReduce Ports

The protocol type of all ports in the table is TCP.

Parameter	Default Port	Port Description
mapreduce.jobhistory.webapp.port	19888	Web HTTP port of the JobHistory server This port is used for: viewing the web page of the JobHistory server **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
mapreduce.jobhistory.port	10020	Port of the JobHistory server This port is used for: 1. Task data restoration in the MapReduce client 2. Obtaining task report in the Job client **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Parameter	Default Port	Port Description
mapreduce.jobhistory.webapp.https.port	19890	Web HTTPS port of the JobHistory server This port is used to view the web page of the JobHistory server. **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Common Spark Ports

The protocol type of all ports in the table is TCP.

Parameter	Default Port	Port Description
hive.server2.thrift.port	22550	JDBC thrift port This port is used for: Socket communication between Spark2.1.0 CLI/ JDBC client and server **NOTE** If **hive.server2.thrift.port** is occupied, an exception indicating that the port is occupied is reported. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Parameter	Default Port	Port Description
spark.ui.port	4040	Web UI port of JDBC This port is used for: HTTPS/HTTP communication between Web requests and the JDBC Server Web UI server **NOTE** The system verifies the port configuration. If the port is invalid, the value of the port plus 1 is used till the calculated value is valid. (A maximum number of 16 attempts are allowed. The number of attempts is specified by **spark.port.maxRetries**.) ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
spark.history.ui.port	18080	JobHistory Web UI port This port is used for: HTTPS/HTTP communication between Web requests and Spark2.1.0 History Server **NOTE** The system verifies the port configuration. If the port is invalid, the value of the port plus 1 is used till the calculated value is valid. (A maximum number of 16 attempts are allowed. The number of attempts is specified by **spark.port.maxRetries**.) ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Common Storm Ports

The protocol type of all ports in the table is TCP.

Parameter	Default Port	Port Description
nimbus.thrift.port	6627	Port for Nimbus to provide thrift services

Parameter	Default Port	Port Description
supervisor.slots.ports	6700,6701,6702,6703	Port for receiving service requests that are forwarded from other servers
logviewer.https.port	29248	Port for LogViewer to provide HTTPS services
ui.https.port	29243	Port for Storm UI to provide HTTPS services (ui.https.port)

Common Yarn Ports

The protocol type of all ports in the table is TCP.

Parameter	Default Port	Port Description
yarn.resourcemanager.webapp.port	8088	Web HTTP port of the ResourceManager service
yarn.resourcemanager.webapp.https.port	8090	Web HTTPS port of the ResourceManager service This port is used to access the Resource Manager web applications in security mode. **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Parameter	Default Port	Port Description
yarn.nodemanager.webapp.port	8042	NodeManager Web HTTP port
yarn.nodemanager.webapp.https.port	8044	NodeManager Web HTTPS port This port is used for: Accessing the NodeManager web application in security mode **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Common ZooKeeper Ports

The protocol type of all ports in the table is TCP.

Parameter	Default Port	Port Description
clientPort	2181	ZooKeeper client port This port is used for: Connection between the ZooKeeper client and server. **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Common Kerberos Ports

The protocol type of all ports in the table is UDP.

Parameter	Default Port	Port Description
kdc_ports	21732	Kerberos server port This port is used for: Performing Kerberos authentication for components. This parameter may be used during the configuration of mutual trust between clusters. **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Common OpenTSDB Ports

The protocol type of the port in the table is TCP.

Parameter	Default Port	Port Description
tsd.network.port	4242	Web UI port of OpenTSDB This port is used for: HTTPS/HTTP communication between web requests and the OpenTSDB UI server

Common Tez Ports

The protocol type of the port in the table is TCP.

Parameter	Default Port	Port Description
tez.ui.port	28888	Web UI port of Tez

Common KafkaManager Ports

The protocol type of the port in the table is TCP.

Parameter	Default Port	Port Description
kafka_manager_port	9099	Web UI port of KafkaManager

Common Presto Ports

The protocol type of the port in the table is TCP.

Parameter	Default Port	Port Description
http-server.http.port	7520	HTTP port for Presto coordinator to provide services to external systems
http-server.https.port	7521	HTTPS port for Presto coordinator to provide services to external systems
http-server.http.port	7530	HTTP port for Presto worker to provide services to external systems
http-server.https.port	7531	HTTPS port for Presto worker to provide services to external systems

Common Flink Ports

The protocol type of the port in the table is TCP.

Parameter	Default Port	Port Description
jobmanager.web.port	32261-32325	Web UI port of Flink This port is used for: HTTP/HTTPS communication between the client web requests and Flink server

Common ClickHouse Ports

The protocol type of the port in the table is TCP.

Parameter	Default Port	Port Description
tcp_port	9000	TCP port for accessing the service client
http_port	8123	HTTP port for accessing the service client
https_port	8443	HTTPS port for accessing the service client
tcp_port_secure	9440	TCP With SSL port for accessing the service client. This port is enabled only in security mode by default.

Common Impala Ports

The protocol type of the port in the table is TCP.

Parameter	Default Port	Port Description
--beeswax_port	21000	Port for impala-shell communication
--hs2_port	21050	Port for Impala application communication
--hs2_http_port	28000	Port used by Impala to provide the HiveServer2 protocol for external systems

7.3 Access Through Direct Connect

MRS allows you to access MRS clusters using Direct Connect. Direct Connect is a high-speed, low-latency, stable, and secure dedicated network connection that connects your local data center to an online cloud VPC. It extends online cloud services and existing IT facilities to build a flexible, scalable hybrid cloud computing environment.

Prerequisites

Direct Connect is available, and the connection between the local data center and the online VPC has been established.

Accessing an MRS Cluster Using Direct Connect

Step 1 Log in to the MRS console.

Step 2 Click the name of the cluster to enter its details page.

Step 3 On the **Dashboard** tab page of the cluster details page, click **Access Manager** next to **MRS Manager**.

Step 4 Set **Access Mode** to **Direct Connect** and select **I confirm that the network between the local PC and the floating IP address is connected and that MRS Manager is accessible using the Direct Connect connection**.

The floating IP address is automatically allocated by MRS to access MRS Manager. Before using Direct Connect to access MRS Manager, ensure that the connection between the local data center and the online VPC has been established.

Step 5 Click **OK**. The MRS Manager login page is displayed. Enter the username **admin** and the password set during cluster creation.

----End

Switching the MRS Manager Access Mode

To facilitate user operations, the browser cache records the selected Manager access mode. To change the access mode, perform the following steps:

Step 1 Log in to the MRS console.

Step 2 Click the name of the cluster to enter its details page.

Step 3 On the **Dashboard** tab page of the cluster details page, click ⇄ next to **MRS Manager**.

Step 4 On the displayed page, set **Access Mode**.

- To change **EIP** to **Direct Connect**, ensure that the network for direct connections is available, set **Access Mode** to **Direct Connect**, and select **I confirm that the network between the local PC and the floating IP address is connected and that MRS Manager is accessible using the Direct Connect connection**. Click **OK**.
- To change **Direct Connect** to **EIP**, set **Access Mode** to **EIP** and configure the EIP by referring to **Accessing FusionInsight Manager Using an EIP**. If a public IP address has been configured for the cluster, click **OK** to access MRS Manager using an EIP.

----End

7.4 EIP-based Access

You can bind an EIP to a cluster to access the web UIs of the open-source components managed in the MRS cluster. This method is simple and easy to use and is recommended for accessing the web UIs of the open-source components.

Binding an EIP to a Cluster and Adding a Security Group Rule

1. On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users. After the IAM users are synchronized, the **Components** tab is available.
2. Click **Access Manager** on the right of **MRS Manager**.
3. The page for accessing MRS Manager is displayed. Bind an EIP and add a security group rule. Perform the following operations only when you access the web UIs of the open-source components of the cluster for the first time.
    a. Select an available EIP from the EIP drop-down list to bind it. If there is no available EIP, click **Manage EIP** to create an EIP. If an EIP has been bound during cluster creation, skip this step.
    b. Select the security group to which the security group rule to be added belongs. The security group is configured when the group is created.
    c. Add a security group rule. By default, your public IP address used for accessing port 9022 is filled in the rule. If you want to view, modify, or delete a security group rule, click **Manage Security Group Rule**.

        NOTE

        - It is normal that the automatically generated public IP address is different from the local IP address and no action is required.
        - If port 9022 is a Knox port, you need to enable the permission of port 9022 to access Knox for accessing MRS components.

    d. Select the checkbox stating that **I confirm that xx.xx.xx.xx is a trusted public IP address and MRS Manager can be accessed using this IP address**.
    e. Click **OK**. The login page is displayed. Enter the username **admin** and the password set during cluster creation.
4. Log in to FusionInsight Manager and choose **Cluster** > **Services** > **HDFS**. On the displayed page, click **NameNode(***Host name***, active)** to access the HDFS web UI. The HDFS NameNode is used as an example. For details about the web UIs of other components, see **Web UIs of Open Source Components**.

7.5 Access Using a Windows ECS

MRS allows you to access the web UIs of open-source components through a Windows ECS. This method is complex and is recommended for MRS clusters that do not support the EIP function.

Step 1 On the MRS management console, click **Clusters**.

Step 2 On the **Active Clusters** page, click the name of the specified cluster.

On the cluster details page, record the **AZ**, **VPC**, **Cluster Manager IP Address**, and **Security Group** of the cluster.

NOTE

To obtain the cluster manager IP address, remotely log in to the Master2 node, and run the **ifconfig** command. In the command output, **eth0:wsom** indicates the cluster manager IP address. Record the value of **inet**. If the cluster manager IP address cannot be queried on the Master2 node, switch to the Master1 node to query and record the cluster manager IP address. If there is only one Master node, query and record the cluster manager IP address of the Master node.

Step 3 On the ECS management console, create an ECS.

- The **AZ**, **VPC**, and **Security Group** of the ECS must be the same as those of the cluster to be accessed.
- Select a Windows public image. For example, select the .
- For details about other configuration parameters, see **Elastic Cloud Server > User Guide > Getting Started > Creating and Logging In to a Windows ECS**.

NOTE

If the security group of the ECS is different from **Security Group** of the MRS cluster, you can modify the configuration using either of the following methods:

- Change the security group of the ECS to the security group of the MRS cluster. For details, see **Elastic Cloud Server** > **User Guide** > **Security Group** > **Changing a Security Group**.
- Add two security group rules to the security groups of the Master and Core nodes to enable the ECS to access the cluster. Set **Protocol** to **TCP** and **ports** of the two security group rules to **28443** and **20009**, respectively. For details, see **Virtual Private Cloud** > **User Guide** > **Security** > **Security Group** > **Adding a Security Group Rule**.

Step 4 On the VPC management console, apply for an EIP and bind it to the ECS.

For details, see **Virtual Private Cloud** > **User Guide** > **Elastic IP** > **Assigning an EIP and Binding It to an ECS**.

Step 5 Log in to the ECS.

The Windows system account, password, EIP, and the security group rules are required for logging in to the ECS. For details, see **Elastic Cloud Server > User Guide > Instances > Logging In to a Windows ECS**.

Step 6 On the Windows remote desktop, use your browser to access Manager.

For example, you can use Internet Explorer 11 in the Windows 2012 OS.

The MRS Manager access address is in the format of **https://*Cluster Manager IP Address*:28443/web**. Enter the name and password of the MRS cluster user, for example, user **admin**.

NOTE

- To obtain the cluster manager IP address, remotely log in to the Master2 node, and run the **ifconfig** command. In the command output, **eth0:wsom** indicates the cluster manager IP address. Record the value of **inet**. If the cluster manager IP address cannot be queried on the Master2 node, switch to the Master1 node to query and record the cluster manager IP address. If there is only one Master node, query and record the cluster manager IP address of the Master node.
- If you access MRS Manager with other MRS cluster usernames, change the password upon your first access. The new password must meet the requirements of the current password complexity policies.
- By default, a user is locked after inputting an incorrect password five consecutive times. The user is automatically unlocked after 5 minutes.

Step 7 Visit the web UIs of the open-source components by referring to the addresses listed in **Web UIs of Open Source Components**.

----End

Related Tasks

Configuring the Mapping Between Cluster Node Names and IP Addresses

Step 1 Log in to MRS Manager, and choose **Host Management**.

Record the host names and management IP addresses of all nodes in the cluster.

Step 2 In the work environment, use Notepad to open the **hosts** file and add the mapping between node names and IP addresses to the file.

Fill in one row for each mapping relationship, as shown in the following figure.

```
192.168.4.127 node-core-Jh3ER
192.168.4.225 node-master2-PaWVE
192.168.4.19 node-core-mtZ81
192.168.4.33 node-master1-zbYN8
192.168.4.233 node-core-7KoGY
```

Save the modifications.

----End

7.6 Creating an SSH Channel for Connecting to an MRS Cluster and Configuring the Browser

Scenario

Users and an MRS cluster are in different networks. As a result, an SSH channel needs to be created to send users' requests for accessing websites to the MRS cluster and dynamically forward them to the target websites.

The MAC system does not support this function. For details about how to access MRS, see **EIP-based Access**.

Prerequisites

- You have prepared an SSH client for creating the SSH channel, for example, the Git open-source SSH client. You have downloaded and installed the client.
- You have created a cluster and prepared a key file in PEM format or obtained the password used during cluster creation.
- Users can access the Internet on the local PC.

Procedure

Step 1 Log in to the MRS management console and choose **Clusters** > **Active Clusters**.

Step 2 Click the specified MRS cluster name.

Record the security group of the cluster.

Step 3 Add an inbound rule to the security group of the Master node to allow data access to the IP address of the MRS cluster through port **22**.

For details, see **Virtual Private Cloud** > **User Guide** > **Security** > **Security Group** > **Adding a Security Group Rule**.

Step 4 Query the primary management node of the cluster. For details, see **Determining Active and Standby Management Nodes of Manager**.

Step 5 Bind an elastic IP address to the primary management node.

For details, see **Virtual Private Cloud** > **User Guide** > **Elastic IP** > **Assigning an EIP and Binding It to an ECS**.

Step 6 Start Git Bash locally and run the following command to log in to the active management node of the cluster: **ssh root@*Elastic IP address*** or **ssh -i *Path of the key file* root@*Elastic IP address***.

Step 7 Run the following command to view data forwarding configurations:

cat /etc/sysctl.conf | grep net.ipv4.ip_forward

- If **net.ipv4.ip_forward=1** is displayed, the forwarding function has been configured. Go to **Step 9**.
- If **net.ipv4.ip_forward=0** is displayed, the forwarding function has not been configured. Go to **Step 8**.
- If **net.ipv4.ip_forward** fails to be queried, this parameter has not been configured. Run the following command and then go to **Step 9**:

  echo "net.ipv4.ip_forward = 1" >> /etc/sysctl.conf

Step 8 Modify forwarding configurations on the node.

1. Run the following command to switch to user **root**:

   sudo su - root

2. Run the following commands to modify forwarding configurations:

   echo 1 > /proc/sys/net/ipv4/ip_forward

   sed -i "s/net.ipv4.ip_forward=0/net.ipv4.ip_forward = 1/g" /etc/sysctl.conf

   sysctl -w net.ipv4.ip_forward=1

3. Run the following command to modify the **sshd** configuration file:

vi /etc/ssh/sshd_config

Press **I** to enter the edit mode. Locate **AllowTcpForwarding** and **GatewayPorts** and delete comment tags. Modify them as follows. Save the changes and exit.

```
AllowTcpForwarding yes
GatewayPorts yes
```

4. Run the following command to restart the sshd service:

   service sshd restart

Step 9 Run the following command to view the floating IP address:

ifconfig

In the command output, **eth0:FI_HUE** indicates the floating IP address of Hue and **eth0:wsom** specifies the floating IP address of Manager. Record the value of **inet**.

Run the **exit** command to exit.

Step 10 Run the following command on the local PC to create an SSH channel supporting dynamic port forwarding:

ssh -i ***Path of the key file*** **-v -ND** ***Local port*** **root@*****Elastic IP address*** or ***ssh -v -ND Local port root@Elastic IP address***. After running the command, enter the password you set when you create the cluster.

In the command, set **Local port** to the user's local port that is not occupied. Port **8157** is recommended.

After the SSH channel is created, add **-D** to the command and run the command to start the dynamic port forwarding function. By default, the dynamic port forwarding function enables a SOCKS proxy process and monitors the user's local port. Port data will be forwarded to the primary management node using the SSH channel.

Step 11 Run the following command to configure the browser proxy.

1. Go to the Google Chrome client installation directory on the local PC.
2. Press **Shift** and right-click the blank area, choose **Open Command Window Here** and enter the following command:

   chrome --proxy-server="socks5://localhost:8157" --host-resolver-rules="MAP * 0.0.0.0 , EXCLUDE localhost" --user-data-dir=c:/tmppath --proxy-bypass-list="*google*com,*gstatic.com,*gvt*.com,*:80"

   NOTE

   – In the preceding command, **8157** is the local proxy port configured in **Step 10**.

   – If the local OS is Windows 10, start the Windows OS, click **Start** and enter **cmd**. In the displayed CLI, run the command in **Step 11.2**. If this method fails, click **Start**, enter the command in the search box, and run the command in **Step 11.2**.

Step 12 In the address box of the browser, enter the address for accessing Manager.

Address format: **https://***Floating IP address of FusionInsight Manager***:28443/web**

The username and password of the MRS cluster need to be entered for accessing clusters with Kerberos authentication enabled, for example, user **admin**. They are not required for accessing clusters with Kerberos authentication disabled.

When accessing Manager for the first time, you must add the address to the trusted site list.

Step 13 Prepare the website access address.

1. Obtain the website address format and the role instance according to **Web UIs**.
2. Click **Services**.
3. Click the specified service name, for example, HDFS.
4. Click **Instance** and view **Service IP Address** of **NameNode(Active)**.

Step 14 In the address bar of the browser, enter the website address to access it.

Step 15 When logging out of the website, terminate and close the SSH tunnel.

----End

8 Accessing Manager

8.1 Accessing FusionInsight Manager (MRS 3.*x* or Later)

Scenario

In MRS 3.*x* or later, FusionInsight Manager is used to monitor, configure, and manage clusters. After the cluster is installed, you can use the account to log in to FusionInsight Manager.

NOTE

If you cannot log in to the WebUI of the component, access FusionInsight Manager by referring to **Accessing FusionInsight Manager from an ECS**.

Accessing FusionInsight Manager Using EIP

Step 1 Log in to the MRS management console.

Step 2 In the navigation pane, choose **Clusters** > **Active Clusters**. Click the target cluster name to access the cluster details page.

Step 3 Click **Manager** next to **MRS Manager**. In the displayed dialog box, configure the EIP information.

1. If no EIP is bound during MRS cluster creation, select an available EIP from the drop-down list on the right of **IEP**. If you have bound an EIP when creating a cluster, go to **Step 3.2**.

   NOTE

   If no EIP is available, click **Manage EIP** to create one. Then, select the created EIP from the drop-down list on the right of **EIP**.

2. Select the security group to which the security group rule to be added belongs. The security group is configured when the cluster is created.
3. Add a security group rule. By default, the filled-in rule is used to access the EIP. To enable multiple IP address segments to access Manager, see steps **Step**

6 to Step 9. If you want to view, modify, or delete a security group rule, click **Manage Security Group Rule**.

4. Select the information to be confirmed and click **OK**.

Step 4 Click **OK**. The Manager login page is displayed.

Step 5 Enter the default username **admin** and the password set during cluster creation, and click **Log In**. The Manager page is displayed.

Step 6 On the MRS management console, choose **Clusters** > **Active Clusters**. Click the target cluster name to access the cluster details page.

NOTE

To grant other users the permission to access Manager, perform **Step 6** to **Step 9** to add the users' public IP addresses to the trusted IP address range.

Step 7 Click **Add Security Group Rule** on the right of **EIP**.

Step 8 On the **Add Security Group Rule** page, add the IP address segment for users to access the public network and select **I confirm that *public network IP/port* is a trusted public IP address. I understand that using 0.0.0.0/0. poses security risks**.

By default, the IP address used for accessing the public network is filled. You can change the IP address segment as required. To enable multiple IP address segments, repeat steps **Step 6** to **Step 9**. If you want to view, modify, or delete a security group rule, click **Manage Security Group Rule**.

Step 9 Click **OK**.

----End

Accessing FusionInsight Manager from an ECS

Step 1 On the MRS management console, click **Clusters**.

Step 2 On the **Active Clusters** page, click the name of the specified cluster.

Record the **AZ**, **VPC**, **MRS ManagerSecurity Group** of the cluster.

Step 3 On the homepage of the management console, choose **Service List** > **Elastic Cloud Server** to switch to the ECS management console and create an ECS.

- The **AZ**, **VPC**, and **Security Group** of the ECS must be the same as those of the cluster to be accessed.
- Select a Windows public image. For example, a standard image **Windows Server 2012 R2 Standard 64bit(40GB)**.
- For details about other configuration parameters, see **Elastic Cloud Server > User Guide > Getting Started > Creating and Logging In to a Windows ECS**.

NOTE

If the security group of the ECS is different from **Default Security Group** of the Master node, you can modify the configuration using either of the following methods:

- Change the security group of the ECS to the default security group of the Master node. For details, see **Elastic Cloud Server** > **User Guide** > **Security Group** > **Changing a Security Group**.
- Add two security group rules to the security groups of the Master and Core nodes to enable the ECS to access the cluster. Set **Protocol** to **TCP**, **Ports** of the two security group rules to **28443** and **20009**, respectively. For details, see .

Step 4 On the VPC management console, apply for an EIP and bind it to the ECS.

For details, see **Virtual Private Cloud** > **User Guide** > **Elastic IP** > **Assigning an EIP and Binding It to an ECS**.

Step 5 Log in to the ECS.

The Windows system account, password, EIP, and the security group rules are required for logging in to the ECS. For details, see **Elastic Cloud Server > User Guide > Instances > Logging In to a Windows ECS**.

Step 6 On the Windows remote desktop, use your browser to access Manager.

For example, you can use Internet Explorer 11 in the Windows 2012 OS.

The address for accessing Manager is the address of the **MRS Manager** page. Enter the name and password of the cluster user, for example, user **admin**.

NOTE

- If you access Manager with other cluster usernames, change the password upon your first access. The new password must meet the requirements of the current password complexity policies. For details, contact the MRS cluster administrator.
- By default, a user is locked after inputting an incorrect password five consecutive times. The user is automatically unlocked after 5 minutes.

Step 7 Log out of FusionInsight Manager. To log out of Manager, move the cursor to ▼ in the upper right corner and click **Log Out**.

----End

8.2 Accessing MRS Manager MRS 2.x or Earlier)

Scenario

MRS uses FusionInsight Manager to monitor, configure, and manage clusters. You can access FusionInsight Manager by clicking **Access Manager** on the **Dashboard** tab page of your MRS cluster and entering username **admin** and the password configured during cluster creation on the login page that is displayed.

Accessing FusionInsight Manager Using an EIP

Step 1 Log in to the MRS management console.

Step 2 In the navigation pane, choose **Clusters** > **Active Clusters**. Click the target cluster name to access the cluster details page.

Step 3 Click **Access Manager** next to **MRS Manager**. In the displayed dialog box, set **Access Mode** to **EIP**. For details about **Direct Connect**, see **Access Through Direct Connect**.

1. If no EIP is bound during MRS cluster creation, select an available EIP from the drop-down list on the right of **IEP**. If you have bound an EIP when creating a cluster, go to **Step 3.2**.

   NOTE

   If no EIP is available, click **Manage EIP** to create one. Then, select the created EIP from the drop-down list on the right of **EIP**.

2. Select the security group to which the security group rule to be added belongs. The security group is configured when the cluster is created.
3. Add a security group rule. By default, your public IP address used for accessing port 9022 is filled in the rule. To enable multiple IP address segments to access MRS Manager, see **Step 6** to **Step 9**. If you want to view, modify, or delete a security group rule, click **Manage Security Group Rule**.

   NOTE

   - It is normal that the automatically generated public IP address is different from the local IP address and no action is required.
   - If port 9022 is a Knox port, you need to enable the permission of port 9022 to access Knox for accessing MRS Manager.

4. Select the checkbox stating that **I confirm that xx.xx.xx.xx is a trusted public IP address and MRS Manager can be accessed using this IP address**.

Step 4 Click **OK**. The MRS Manager login page is displayed.

Step 5 Enter the default username **admin** and the password set during cluster creation, and click **Log In**. The MRS Manager page is displayed.

Step 6 On the MRS management console, choose **Clusters** > **Active Clusters**, and click the target cluster name to access the cluster details page.

NOTE

To assign MRS Manager access permissions to other users, follow instructions from **Step 6** to **Step 9** to add the users' public IP addresses to the trusted range.

Step 7 Click **Add Security Group Rule** on the right of **EIP**.

Step 8 On the **Add Security Group Rule** page, add the IP address segment for users to access the public network and select **I confirm that the authorized object is a trusted public IP address range. Do not use 0.0.0.0/0. Otherwise, security risks may arise**.

By default, the IP address used for accessing the public network is filled. You can change the IP address segment as required. To enable multiple IP address segments, repeat steps **Step 6** to **Step 9**. If you want to view, modify, or delete a security group rule, click **Manage Security Group Rule**.

Step 9 Click **OK**.

----End

Accessing MRS Manager Using an ECS

Step 1 On the MRS management console, click **Clusters**.

Step 2 On the **Active Clusters** page, click the name of the specified cluster.

Record the **AZ**, **VPC**, and **Security Group** of the cluster.

Step 3 On the ECS management console, create an ECS.

- The **AZ**, **VPC**, and **Security Group** of the ECS must be the same as those of the cluster to be accessed.
- Select a Windows public image. For example, select the .
- For details about other configuration parameters, see .

NOTE

If the security group of the ECS is different from **Default Security Group** of the MRS cluster, you can modify the configuration using either of the following methods:

- Change the default security group of the ECS to the security group of the MRS cluster. For details, see .
- Add two security group rules to the security groups of the Master and Core nodes to enable the ECS to access the cluster. Set **Protocol** to **TCP** and **ports** of the two security group rules to **28443** and **20009**, respectively. For details, see .

Step 4 On the VPC management console, apply for an EIP and bind it to the ECS.

For details, see **Virtual Private Cloud** > **User Guide** > **Elastic IP** > **Assigning an EIP and Binding It to an ECS**.

Step 5 Log in to the ECS.

The Windows system account, password, EIP, and the security group rules are required for logging in to the ECS. For details, see .

Step 6 On the Windows remote desktop, use your browser to access Manager.

For example, you can use Internet Explorer 11 in the Windows 2012 OS.

The Manager access address is in the format of **https://***Cluster Manager IP Address***:28443/web**. Enter the name and password of the MRS cluster user, for example, user **admin**.

NOTE

- To obtain the cluster manager IP address, remotely log in to the Master2 node, and run the **ifconfig** command. In the command output, **eth0:wsom** indicates the cluster manager IP address. Record the value of **inet**. If the cluster manager IP address cannot be queried on the Master2 node, switch to the Master1 node to query and record the cluster manager IP address. If there is only one Master node, query and record the cluster manager IP address of the Master node.
- If you access MRS Manager with other MRS cluster usernames, change the password upon your first access. The new password must meet the requirements of the current password complexity policies.
- By default, a user is locked after inputting an incorrect password five consecutive times. The user is automatically unlocked after 5 minutes.

Step 7 Log out of FusionInsight Manager. To log out of Manager, move the cursor to in the upper right corner and click **Log Out**.

----End

Changing an EIP for a Cluster

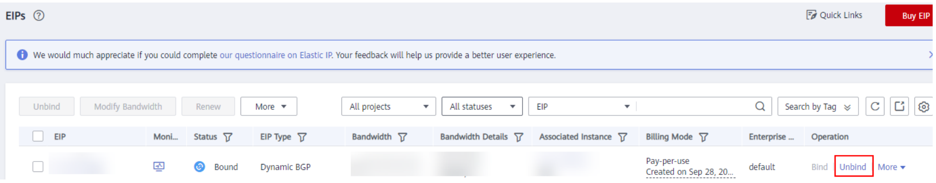
Step 1 On the MRS management console, choose **Clusters** > **Active Clusters**, and click the target cluster name to access the cluster details page.

Step 2 View EIPs

Step 3 Log in to the VPC management console.

Step 4 Choose **Elastic IP and Bandwidth** > **EIPs**.

Step 5 Search for the EIP bound to the MRS cluster and click **Unbind** in the **Operation** column to unbind the EIP from the MRS cluster.



We would much appreciate if you could complete our questionnaire on Elastic IP. Your feedback will help us provide a better user experience.

EIP	Moni...	Status	EIP Type	Bandwidth	Bandwidth Details	Associated Instance	Billing Mode	Enterprise ...	Operation
		Bound	Dynamic BGP				Pay-per-use Created on Sep 28, 20...	default	Bind Unbind More

Step 6 Log in to the MRS management console, choose **Clusters** > **Active Clusters**, and click the target cluster name to access the cluster details page.

EIP on the cluster details page is displayed as **Unbound**.

Step 7 Click **Access Manager** next to **MRS Manager**. In the displayed dialog box, set **Access Mode** to **EIP**.

Step 8 Select a new EIP from the EIP drop-down list and configure other parameters. For details, see **Accessing FusionInsight Manager Using an EIP**.

----End

Granting the Permission to Access MRS Manager to Other Users

Step 1 On the MRS management console, choose **Clusters** > **Active Clusters**, and click the target cluster name to access the cluster details page.

Step 2 Click **Add Security Group Rule** on the right of **EIP**.

Step 3 On the **Add Security Group Rule** page, add the IP address segment for users to access the public network and select **I confirm that the authorized object is a trusted public IP address range. Do not use 0.0.0.0/0. Otherwise, security risks may arise**.

By default, the IP address used for accessing the public network is filled. You can change the IP address segment as required. To enable multiple IP address segments, repeat steps **Step 1** to **Step 4**. If you want to view, modify, or delete a security group rule, click **Manage Security Group Rule**.

Step 4 Click **OK**.

----End

9 FusionInsight Manager Operation Guide (Applicable to 3.x)

9.1 Getting Started

9.1.1 FusionInsight Manager Introduction

Overview

MRS allows you to manage and analyze massive amounts of structured and unstructured data for rapid data mining. Open source components have complex structures and therefore they are difficult to install, configure, and manage. FusionInsight Manager is a unified enterprise-level cluster management platform that provides:

- **Cluster monitoring**: enables you to quickly learn the running status of hosts and services.
- **Graphical metric monitoring and customization**: enable you to obtain key system information in a timely manner.
- **Service property configuration**: allows you to configure service properties based on the performance requirements of your services.
- **Cluster, service, and role instance operations**: allow you to start or stop services and clusters with just a few clicks.
- **Rights management and audit**: allow you to configure the access control and manage operation logs.

Introduction to the FusionInsight Manager GUI

FusionInsight Manager provides a unified cluster management platform, facilitating rapid and easy O&M for clusters.

The upper part of the page is the operation bar, the middle part is the display area, and the bottom part is the taskbar.

Table 9-1 describes the functions of each module on the top menu bar.

Table 9-1 Functions of each module on the top menu bar

Module	Function
Homepage	Displays key monitoring metrics of clusters and host statuses in column charts, line charts, and tables. You can customize a dashboard for key monitoring metrics and drag them onto any positions on the UI. The **Summary** tab page supports automatic data update. For details, see **Homepage**.
Cluster	Provides guidance on how to monitor, operate, and configure services in a cluster, helping you manage services in a unified manner. For details, see **Cluster**.
Hosts	Provides guidance on how to monitor and operate hosts, helping you manage hosts in a unified manner. For details, see **Hosts**.
O&M	Provides guidance on how to query and handle alarms, helping you identify and rectify product faults and potential risks in a timely manner to ensure smooth system running. For details, see **O&M**.
Audit	Allows you to query and export audit logs, and view all user activities and operations. For details, see **Audit**.
Tenant Resources	Provides a unified tenant management platform. For details, see **Tenant Resources**.
System	Provides system management settings of FusionInsight Manager, such as user permission settings. For details, see **System**.

9.1.2 Querying the FusionInsight Manager Version

By viewing the FusionInsight Manager version, you can prepare for system upgrade and routine maintenance.

- Using the GUI:

  Log in to FusionInsight Manager. On the home page, click ? in the upper right corner and choose **About** from the drop-down list. In the dialog box that is displayed, view the FusionInsight Manager version.

- Using the CLI

  a. Log in to the FusionInsight Manager active management node as user **root**.

  b. Run the following commands to check the version and platform information of FusionInsight Manager:

  ```
  su - omm
  cd ${BIGDATA_HOME}/om-server/om/sbin/pack
  ./queryManager.sh
  ```

The following information is displayed:

```
  Version          Package                                          Cputype
  ***           FusionInsight_Manager_***                                x86_64
```

NOTE

******* indicates the version number. Replace it with the actual version number.

9.1.3 Logging In to FusionInsight Manager

Scenario

Log in to FusionInsight Manager using an account.

Procedure

Step 1 Obtain the URL for logging in to FusionInsight Manager.

Step 2 On login page, enter the username and password.

Step 3 Change the password upon your first login.

The password must:

- Contain 8 to 64 characters.
- Contain at least four types of the following characters: uppercase letters, lowercase letters, digits, spaces, and special characters (`~!@#$%^&*()-_=+|[{}];',<.>/\?).
- Be different from the username or the username spelled backwards.
- Be different from the current password.

Step 4 Move the cursor over ▼ in the upper right corner of home page, and choose **Logout** from the drop-down list. In the dialog box that is displayed, click **OK** to log out of the current user.

----End

9.1.4 Logging In to the Management Node

Scenario

Some O&M operation scripts and commands need to be run or can be run only on the active management node. You can identify and log in to the active or standby management node based on the following operations.

Checking and Logging In to the Active and Standby Management Nodes

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **OMS**.

In the **Basic Information** area, **Current Active** indicates the host name of the active management node, and **Current Standby** indicates the host name of the standby management node.

Click a host name to go to the host details page. On the host details page, record the IP address of the host.

Step 3 Log in to the active or standby management node as user **root**.

----End

Identifying the Active and Standby Management Nodes by Running Scripts and Logging In to Them

Step 1 Log in to any node where FusionInsight Manager is installed as user **root**.

Step 2 Run the following command to identify the active and standby management nodes:

su - omm

sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh

In the command output, the node whose **HAActive** is **active** is the active management node (Master1), and the node whose **HAActive** is **standby** is the standby node (Master2).

```
HAMode
double
NodeName            HostName       HAVersion          StartTime              HAActive
HAAllResOK          HARunPhase
192-168-0-30        Master1        V100R001C01        2021-09-01 07:12:05      active
normal              Actived
192-168-0-24        Master2        V100R001C01        2021-09-01 07:14:02      standby
normal              Deactived
```

Step 3 Run the following command to obtain the IP addresses of the active and standby management nodes:

cat /etc/hosts

Example IP addresses of the active and standby management nodes:

```
127.0.0.1       localhost
192.168.0.30    Master1
192.168.0.24    Master2
```

Step 4 Log in to the active or standby management node as user **root**.

----End

9.2 Homepage

9.2.1 Overview

After you log in to FusionInsight Manager, **Homepage** is displayed by default. On this page, the **Summary** tab displays the service statuses and monitoring status reports of each cluster, and the **Alarm Analysis** tab displays the statistics and analysis of top alarms.

- On the right of the home page, you can view the number of alarms of different severities, number of running tasks, current user, and help information.

- Click ▤ to view the task name, cluster, status, progress, start time, and end time of the last 100 operation tasks in **Task Management Center**.

  NOTE

  For a start, stop, restart, or rolling restart task, you can abort it by clicking the task name in the task list, clicking **Abort**, and then entering the system administrator password in the dialog box that is displayed. An aborted task is no longer executed.

- Click ? to obtain help information.

Table 9-2 Help information

Item	Description
About	Provides the FusionInsight Manager version information.

- The taskbar at the bottom of the home page displays the language options of FusionInsight Manager and the current cluster time and time zone information. You can switch the system language as needed.

Service Status Preview Area

The number of hosts available in and the number of services installed in each cluster are displayed on the left of the homepage. You can click ︾ to expand all service information of the cluster and view the status and alarms of each service.

Click ⋯ to perform basic O&M management operations on the current cluster. For details, see **Table 9-3**.

The 🟢 icon on the left of each service name indicates that the service is running properly; the 🔴 icon indicates that the current service fails to start; and the ● icon indicates that the current service is not started.

You can also check whether alarms have been generated for the service on the right of the service name. If alarms have been generated, the alarm severities and the number of alarms are displayed.

For components that support multiple services, if multiple services have been installed in the same cluster, the number of installed services is displayed on the right of each component.

The ⟳ icon displayed on the right of the service name indicates that the service configuration has expired.

Monitoring Status Report Area

The chart area is on the right of the homepage, which displays key monitoring metric reports, such as the status of all hosts in the cluster, host CPU usage, and host memory usage. You can customize monitoring reports to display in this area. For details about how to manage monitoring metrics, see **Managing Monitoring Metric Reports**.

You can view the data source of a monitoring chart in the lower left corner of the chart. You can zoom in on a monitoring report to view chart values more clearly or close the monitoring report.

Alarm Analysis

On the **Alarm Analysis** tab page, you can view the **Top 20 Alarms** table and **Analysis on Top 3 Alarms** chart. You can click an alarm name in the **Top 20 Alarms** table to view analysis information of this alarm only. Alarm analysis allows you to view top alarms and their occurrence time so you can handle alarms accordingly, improving system stability.

9.2.2 Managing Monitoring Metric Reports

Scenario

On FusionInsight Manager, you can customize monitoring items to display on the homepage and export monitoring data.

NOTE

The interval on the horizontal axis of the chart varies depending on the time period you specify. Data monitoring rules are as follows:

- **0 to 25 hours**: The interval is 5 minutes. The cluster must have been installed for at least 10 minutes, and monitoring data of a maximum of 15 days is saved.
- **25 to 150 hours**: The interval is 30 minutes. The cluster must have been installed for at least 30 minutes, and monitoring data of a maximum of 3 months is saved.
- **150 to 300 hours**: The interval is 1 hour. The cluster must have been installed for at least 1 hour, and monitoring data of a maximum of 3 months is saved.
- **300 hours to 300 days**: The interval is 1 day. The cluster must have been installed for at least 1 day, and monitoring data of a maximum of 6 months is saved.
- **Over 300 days**: The interval is 7 days. The cluster must have been installed for more than 7 days, and monitoring data of a maximum of 1 year is saved.
- If the disk usage of the partition where GaussDB resides exceeds 80%, real-time monitoring data and monitoring data whose interval is 5 minutes will be deleted.
- **Storage resources (HDFS) in Tenant Resources (0 to 300 hours)**: The interval is 1 hour. The cluster must have been installed for at least 1 hour, and monitoring data of a maximum of 3 months is saved.

Customizing a Monitoring Metric Report

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Homepage**.

Step 3 In the upper right corner of the chart area, click ▼ and choose **Customize** from the displayed menu.

NOTE

Monitoring data of the past 1 hour is displayed at an interval of 5 minutes. After you enter the **Real-time Monitoring** page, you can view that real-time monitoring data is displayed on the right of the monitoring chart at an interval of 5 minutes.

Step 4 In the left pane of the **Customize Statistics** dialog box, select a resource to monitor.

Step 5 Select one or multiple monitoring metrics in the right pane.

Step 6 Click **OK**.

----End

Exporting All Monitoring Data

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Homepage**.

Step 3 In the upper right corner of the chart area, select a time range to obtain monitoring data, for example, **1w**.

Real-time data is displayed by default, which cannot be exported. You can click ✏ to customize a time range.

Step 4 In the upper right corner of the chart area, click ▼ and choose **Export** from the displayed menu.

----End

Exporting Monitoring Data of a Specified Monitoring Item

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Homepage**.

Step 3 Click ⤡ in the upper right corner of any monitoring report pane in the chart area of the target cluster.

Step 4 Select a time range to obtain monitoring data, for example, **1w**.

Real-time data is displayed by default, which cannot be exported. You can click ✏ to customize a time range.

Step 5 Click **Export**.

----End

9.3 Cluster

9.3.1 Cluster Management

9.3.1.1 Overview

Dashboard

Log in to FusionInsight Manager and choose **Cluster** > *Name of the desired cluster* > **Dashboard** to view the status of the current cluster.

On the **Dashboard** tab page, you can start, stop, perform a rolling restart of, synchronize configurations to, and perform other basic operations on the current cluster.

Table 9-3 Maintenance and management operations

UI Portal	Description
Start	Starts all services in the cluster.
Stop	Stops all services in the cluster.
More > **Restart**	Restarts all services in the cluster.
More > **Rolling-restart Service**	Restarts all services in the cluster one at a time without interrupting workloads. For details about how to perform a rolling restart, see **Performing a Rolling Restart of a Cluster**.
More > **Synchronize Configurations**	Enables new configuration parameters for all services in the cluster.
More > **Restart Configuration-Expired Instances**	Restarts expired instances for all services in the cluster. For details, see **Managing Expired Configurations**.
More > **Health Check**	Performs a health check on the OMS nodes, all services, and the rest nodes in the cluster. There are three types of check items: running status, related alarms, and custom monitoring metrics. The health check results are not always the same as the values of **Running Status** displayed on the GUI. You can export check results by clicking **Export** in the upper left corner of the checklist. If any issues are detected, you can click **View Help** to find a troubleshooting method.
More > **Download Client**	Downloads the default client. For details, see **Downloading the Client**.
More > **Export Installation Template**	Batch exports all installation configurations of the cluster, such as the cluster authentication mode, node information, and service configuration. You can use this function when you need to reinstall the cluster in the same environment.
More > **Export Configurations**	Batch exports configurations of all services in the cluster.
More > **Enter Maintenance Mode** and **More** > **Exit Maintenance Mode**	Enters or exits the cluster maintenance mode.
More > **O&M View**	Allows you to view services or hosts that are in the maintenance mode.

9.3.1.2 Performing a Rolling Restart of a Cluster

Scenario

A rolling restart is batch restarting all services in a cluster after they are modified or upgraded without interrupting workloads.

You can perform a rolling restart of a cluster as needed.

NOTE

- Certain services in a cluster do not support rolling restart. These services are restarted in normal mode during the rolling restart of the cluster. As a result, workloads may be interrupted. So, you need to determine whether to perform this operation as prompted.
- Configurations that must take effect immediately, for example, server port configurations, should be restarted in normal mode.

Impact on the System

A rolling restart takes a longer time and may affect service throughput and performance.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the target cluster* > **Dashboard**. On this tab page, choose **More** > **Rolling-restart Service**.

Step 3 In the dialog box that is displayed, enter the password of the current login user and click **OK**.

Step 4 Configure the parameters based on site requirements.

Table 9-4 Rolling restart parameters

Parameter	Description
Restart only instances with expired configurations in the cluster	Whether to restart only the modified instances in a cluster
Enable rack strategy	Whether to enable the concurrent rack rolling restart strategy. This parameter takes effect only for roles that meet the rack rolling restart strategy. (The roles support rack awareness, and instances of the roles belong to two or more racks.) **NOTE** This parameter is configurable only when a rolling restart is performed on HDFS or YARN.

Parameter	Description
Data Nodes to Be Batch Restarted	Number of instances that are restarted in each batch when the batch rolling restart strategy is used. The default value is **1**. **NOTE** - This parameter is valid only when the batch rolling restart strategy is used and the instance type is DataNode. - This parameter is invalid when the rack strategy is enabled. In this case, the cluster uses the maximum number of instances (20 by default) configured in the rack strategy as the maximum number of instances that are concurrently restarted in a rack. - This parameter is configurable only when a rolling restart is performed on HDFS, HBase, YARN, Kafka, Storm, or Flume. - This parameter for the RegionServer of HBase cannot be manually configured. Instead, it is automatically adjusted based on the number of RegionServer nodes. Specifically, if the number of RegionServer nodes is less than 30, the parameter value is **1**. If the number is greater than or equal to 30 and less than 300, the parameter value is **2**. If the number is greater than or equal to 300, the parameter value is 1% of the number (rounded-down).
Batch Interval	Interval between two batches of instances to be roll-restarted. The default value is **0**.
Decommissioning Timeout Interval	Decommissioning interval for role instances during a rolling restart. The default value is **1800s**. Some roles (such as HiveServer and JDBCServer) stop providing services before the rolling restart. Stopped instances cannot cannot be connected to new clients. Existing connections will be completed after a period of time. An appropriate timeout interval can ensure service continuity. **NOTE** This parameter is configurable only when a rolling restart is performed on Hive or Spark2x.
Batch Fault Tolerance Threshold	Tolerance times when the rolling restart of instances fails to be batch executed. The default value is **0**, which indicates that the rolling restart task ends after any batch of instances fails to restart.

NOTE

Advanced parameters, such as **Data Nodes to Be Batch Restarted**, **Batch Interval**, and **Batch Fault Tolerance Threshold**, should be properly configured based on site requirements. Otherwise, services may be interrupted or cluster performance may be severely affected.

Example:

- If **Data Nodes to Be Batch Restarted** is set to an unnecessarily large value, a large number of instances are restarted concurrently. As a result, services are interrupted or cluster performance is severely affected due to too few working instances.
- If **Batch Fault Tolerance Threshold** is too large, services will be interrupted because a next batch of instances will be restarted after a batch of instances fails to restart.

Step 5 Click **OK**.

----End

9.3.1.3 Managing Expired Configurations

Scenario

If a new configuration needs to be delivered to all services in the cluster, or **Configuration Status** of multiple services changes to **Expired** or **Failed** after a configuration is modified, the configuration parameters of these services are not synchronized and do not take effect. In this case, synchronize the configurations and restart related service instances for the cluster so that the new parameters take effect for all services.

If the configuration of the services in the cluster has been synchronized but do not take effect, you need to restart the instances whose configuration has expired.

Impact on the System

- After synchronizing the cluster configuration, you need to restart the services whose configuration has expired. These services are unavailable during restart.
- The instances whose configuration has expired are unavailable during restart.

Procedure

Synchronize the configuration.

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Dashboard**.

Step 3 On this page, choose **More** > **Synchronize Configuration**.

Step 4 In the dialog box that is displayed, click **OK**.

----End

Restart configuration-expired instances.

Step 1 Choose **More** > **Restart Configuration-Expired Instances**.

Step 2 In the dialog box that is displayed, enter the password of the current login user and click **OK**.

Step 3 In the displayed dialog box, click **OK**.

You can click **View Instance** to open the list of all expired instances and confirm that the instances have been restarted.

----End

9.3.1.4 Downloading the Client

Scenario

Use the default client provided by MRS clusters to manage the cluster, run services, and perform secondary development. Before you use this client, you need to download its software package.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Dashboard**. On the page that is displayed, choose **More** > **Download Client**.

The **Download Cluster Client** dialog box is displayed.

Step 3 Select a client type for **Select Client Type**.

- **Complete Client**: the package contains scripts, compilation files, and configuration files.
- **Configuration Files Only**: the package contains only the client configuration files.

  This type is applicable to application development tasks. For example, after a complete client is downloaded and installed, the cluster administrator modifies the service configuration on FusionInsight Manager, and developers need to update the client configuration files.

NOTE

Set **Select Platform Type** to **x86_64** or **aarch64**. To run the client on x86 nodes, select **x86_64**; to tun the client on TaiShan nodes, select **aarch64**. By default, you should select a client that has the same architecture as your servers.

Step 4 Determine whether to generate a client software package file on the cluster node.

- If yes, select **Save to Path** and click **OK** to generate the client file.

  The generated file is stored in the **/tmp/FusionInsight-Client** directory on the active management node by default. You can also store the client file in other directories, and user **omm** has the read, write, and execute permissions on the directory. If the client file already exists in the path, the existing client file will be replaced.

  After the file is generated, copy the obtained package to another directory, for example, **/opt/Bigdata/hadoopclient**, as user **omm** or client installation user.
- If no, click **OK** to download the client file to the local host.

  The system starts to download the client software package.

Install the downloaded client by referring to **Installing a Client**.

----End

9.3.1.5 Modifying Cluster Attributes

Scenario

View basic cluster attributes on FusionInsight Manager.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Cluster Properties**.

By default, you can view the cluster name, cluster description, product type, cluster ID, authentication mode, creation time, and installed components.

Step 3 Change the cluster name.

1. Click ✎ and enter a new name.

   Enter 2 to 199 characters. Only letters, digits, underscores (_), hyphens (-), and spaces are allowed, and the name cannot start with a space.
2. Click **OK** for the new cluster name to take effect.

Step 4 Modify the cluster description.

1. Click ✎ and enter a new description.

   Enter a maximum of 199 characters. Only letters, digits, commas (,), periods (.), underscores (_), spaces, and newline characters (\n) are allowed.
2. Click **OK** for the new description to take effect.

----End

9.3.1.6 Managing Cluster Configurations

Scenario

FusionInsight Manager allows you to view the changes of service configuration parameters in a cluster with one click, helping you quickly locate faults and improve configuration management efficiency.

You can quickly view all non-default values of each service in the cluster, non-uniform values between instances of the same role, historical records of cluster configuration modifications, and expired parameters in the cluster on the configuration page.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Configurations**.

Step 3 Select an operation page based on the scenario.

- To view all non-default values:
  a. Click **All Non-default Values**. The system displays the parameters whose values are different from the default values configured for each service, role, or instance in the current cluster.

     You can click ↻ next to a parameter value to quickly restore the value to the default one. You can click 🕘 to view the historical modification records of the parameter.

     If there are a large number of parameters to configure, you can filter the parameters in the filter box in the upper right corner of the page or enter keywords in the search box.

  b. To change the values of the parameters, change the values according to the parameter description and click **Save**. In the dialog box that is displayed, click **OK**.
- To view all non-uniform values:
  a. Click **All Non-uniform Values**. The system displays parameters with different role, service, instance group, or instance configurations in the current cluster.

     You can click ⛶ next to a parameter value and view the differences in the dialog box that is displayed.

  b. To change the value of a parameter, click 🧹 to cancel the configuration difference or manually adjust the parameter value, click **OK**, and then click **Save**. In the dialog box that is displayed, click **OK**.
- To check expired configurations:
  a. Click **Expired Configurations**. Expired configuration items in the current cluster are displayed.
  b. You can filter services using the service filter box in the upper part of the page to view expired configurations of different services. Alternatively, you can enter keywords in the search box.
  c. Expired configuration items do not take effect completely. Restart the services or instances whose configurations have expired in a timely manner.
- To view historical configuration records:
  a. Click **Historical Configurations**. The historical configuration change records of the current cluster are displayed. You can view details about parameter value changes, including the service to which the parameter belongs, parameter values before and after the modification, and parameter files.
  b. To restore a configuration change, click **Restore Configuration** in the **Operation** column of the target record. In the dialog box that is displayed, click **OK**.

NOTE

Some configuration items take effect only after the corresponding services are restarted. After the configurations are saved, restart the services or instances whose configurations have expired in a timely manner.

----End

9.3.1.7 Managing Static Service Pools

9.3.1.7.1 Static Service Resources

Overview

A cluster allocates static service resources to services Flume, HBase, HDFS, and YARN. The total volume of computing resources allocated to each service is fixed, and they are static. A tenant can exclusively use or share a service to obtain the resources required for running this service.

Static Service Pool

Static service pools are used to specify service resource configurations.

Static service pools centrally manage resources that can be used by each service.

- Limits the total number of resources that can be used by each services. Specifically, the total number of CPU, I/O, and memory resources can be configured on the nodes where services Flume, HBase, HDFS, and YARN are deployed.
- Isolates the resources of services in a cluster from those of other services. In this way, the load of one service has very limited impact on other services.

Scheduling Mechanism

The time-based dynamic resource scheduling mechanism enables different volumes of static resources to be configured for services at different time, optimizing service running environments and improving the cluster efficiency.

In a complex cluster environment, multiple services share resources in the cluster, but the resource service period of each service may be different.

The following use a bank customer as an example:

- The HBase query service is heavy in the daytime.
- The query service is light, but the Hive analysis service is heavy at night.

If fixed resources are allocated to each service, the following problems may occur:

- The query service cannot obtain sufficient resources while the resources for the analysis service are idle in the daytime.
- The analysis service cannot obtain sufficient resources while the resources for the query service are idle at night.

As a result, the cluster resource utilization is low and the service capability is weak. Resolve the problem in the following ways:

- Sufficient resources need to be configured for HBase in the daytime.
- Sufficient resources need to be configured for Hive at night.

The time-based dynamic scheduling mechanism can efficiently utilize resources and run tasks.

9.3.1.7.2 Configuring Cluster Static Resources

Scenario

You can adjust resource base on FusionInsight Manager and customize resource configuration groups if you need to control service resources used on each node in a cluster or the available CPU or I/O quotas on each node at different time segments.

Impact on the System

- After a static service pool is configured, the configuration status of affected services is displayed as **Expired**. You need to restart the services. Services are unavailable during restart.
- After a static service pool is configured, the maximum number of resources used by each service and role instance cannot exceed the upper limit.

Procedure

Modify the Resource Adjustment Base

Step 1 On FusionInsight Manager, choose **Cluster**, click the name of the target cluster, and choose **Static Service Pool Configurations**.

Step 2 Click **Configuration** in the upper right corner. The page for configuring resource pools is displayed.

Step 3 Change the values of **CPU (%)** and **Memory (%)** in the **System Resource Adjustment Base** area.

Modifying the system resource adjustment base changes the maximum physical CPU and memory usage on nodes by services. If multiple services are deployed on the same node, the maximum physical resource usage of all services cannot exceed the adjusted CPU or memory usage.

Step 4 Click **Next**.

To modify parameters again, click **Previous**.

Modify the Default Resource Configuration Group

Step 5 Click **default**. In the **Configure weight** table, set **CPU LIMIT(%)**, **CPU SHARE(%)**, **I/O(%)**, and **Memory(%)** for each service.

NOTE

- The sum of **CPU LIMIT(%)** and **CPU SHARE(%)** used by all services can exceed 100%.
- The sum of **I/O(%)** used by all services can exceed 100% but cannot be 0.
- The sum of **Memory(%)** used by all services can be greater than, smaller than, or equal to 100%.
- **Memory(%)** cannot take effect dynamically and can only be modified in the default configuration group.
- **CPU LIMIT(%)** is used to configure the ratio of the number of CPU cores that can be used by a service to those can be allocated to related nodes.
- **CPU SHARE(%**) is used to configure the ratio of the time when a service uses a CPU core to the time when other services use the CPU core. That is, the ratio of time when multiple services compete for the same CPU core.

Step 6 Click **Generate detailed configurations based on weight configurations**. FusionInsight Manager generates the actual values of the parameters in the default weight configuration table based on the cluster hardware resources and allocation information.

Step 7 Click **OK**.

In the displayed dialog box, click **OK**.

Add a Customized Resource Configuration Group

Step 8 Determine whether to automatically adjust resource configurations at different time segments.

- If yes, go to **Step 9**.
- If no, use the default configurations, and no further action is required.

Step 9 Click **Configuration**, change the system resource adjustment base values, and click **Next**.

Step 10 Click **Add** to add a resource configuration group.

Step 11 In **Step 1: Scheduling Time**, click **Configuration**.

The page for configuring the time policy is displayed.

Modify the following parameters based on service requirements and click **OK**.

- **Repeat**: If this parameter is selected, the customized resource configuration is applied repeatedly based on the scheduling period. If this parameter is not selected, set the date and time when the configuration of the group of resources can be applied.
- **Repeat Policy**: The available values are **Daily**, **Weekly**, and **Monthly**. This parameter is valid only when **Repeat** is selected.
- **On**: indicates the time period between the start time and end time when the resource configuration is applied. Set a unique time range. If the time range overlaps with that of an existing group of resource configuration, the time range cannot be saved.

NOTE

- The default group of resource configuration takes effect in all undefined time segments.
- The newly added resource group is a parameter set that takes effect dynamically in a specified time range.
- The newly added resource group can be deleted. A maximum of four resource configuration groups that take effect dynamically can be added.
- Select a repetition policy. If the end time is earlier than the start time, the resource configuration ends in the next day by default. For example, if a validity period ranges from 22:00 to 06:00, the customized resource configuration takes effect from 22:00 on the current day to 06:00 on the next day.
- If the repetition policy types of multiple configuration groups are different, the time ranges can overlap. The policy types are listed as follows by priority from low to high: daily, weekly, and monthly. The following is an example. There are two resource configuration groups using the monthly and daily policies, respectively. Their application time ranges in a day overlap as follows: 04:00 to 07:00 and 06:00 to 08:00. In this case, the configuration of the group that uses the monthly policy prevails.
- If the repetition policy types of multiple resource configuration groups are the same, the time ranges of different dates can overlap. For example, if there are two weekly scheduling groups, you can set the same time range on different day for them, such as to 04:00 to 07:00, on Monday and Wednesday, respectively.

Step 12 Modify the resource configuration of each service in **Step 2: Weight Configuration**.

Step 13 Click **Generate detailed configuration**. FusionInsight Manager generates the actual values of the parameters in the default weight configuration table based on the cluster hardware resources and allocation information.

Step 14 Click **OK**.

In the displayed dialog box, click **OK**.

----End

9.3.1.7.3 Viewing Cluster Static Resources

Scenario

The big data management platform can manage and isolate service resources that are not running on YARN using static service resource pools. The system supports time-based automatic adjustment of static service resource pools. This enables the cluster to automatically adjust the parameter values at different periods to ensure more efficient resource utilization.

System administrators can view the monitoring indicators of resources used by each service in the static service pool on FusionInsight Manager. The monitoring indicators are as follows:

- CPU usage of services
- Total disk I/O read rate of services
- Total disk I/O write rate of services
- Total used memory of services

NOTE

After the multi-tenant function is enabled, the CPU, I/O, and memory usage of all HBase instances can be centrally managed.

Procedure

Step 1 On FusionInsight Manager, choose **Cluster**, click the name of the target cluster, and choose **Static Service Pool Configurations**.

Step 2 In the configuration group list, click a configuration group, for example, **default**.

Step 3 Check the system resource adjustment base values.

- **System Resource Adjustment Base** indicates the maximum volume of resources that can be used by each node in the cluster. If a node has only one service, the service exclusively occupies the available resources on the node. If a node has multiple services, all services share the available resources on the node.
- **CPU** indicates the maximum number of CPUs that can be used by services on a node.
- **Memory** indicates the maximum memory that can be used by services on a node.

Step 4 In **Chart**, view the metric data of the cluster service resource usage.

NOTE

- You can click **Add Service to Chart** to add static service resource data of specific services (up to 12 services) to the chart.
- For details about how to manage a chart, see **Managing Monitoring Metric Reports**.

----End

9.3.1.8 Managing Clients

9.3.1.8.1 Managing a Client

Scenario

FusionInsight Manager supports unified management of cluster client installation information. After a user downloads and installs a client, FusionInsight Manager automatically records information about the installed (registered) client to facilitate query and management. In addition, you can manually add or modify the information about clients that are not automatically registered, for example, clients installed in earlier versions.

Procedure

View client information.

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster**, click the name of the desired cluster, and choose **Client Management** to view information about clients installed in the cluster.

You can view the IP address, installation path, component list, registration time, and installation user of the node where the client is located.

When the client is downloaded and installed in the cluster of the latest version, the client information is automatically registered.

Add client information.

Step 3 To manually add information about an installed client, click **Add** and manually add the IP address, installation path, user, platform information, and registration information of the client as prompted.

Step 4 Configure the client information and click **OK**.

Modify client information.

Step 5 Modify information about the manually registered client.

On the **Client Management** page, select the target client and click **Modify**. After modifying the information, click **OK**.

Delete client information.

Step 6 On the **Client Management** page, select the target client and click **Delete**. In the displayed dialog box, click **OK**.

To delete multiple clients, select the all of them and click **Batch Delete**. In the displayed dialog box, click **OK**.

Export client information.

Step 7 On the **Client Management** page, click **Export All** to export information about all registered clients to the local PC.

NOTE

On the **Client Management** page, only components that have clients are displayed in the component list. Therefore, some components that do not have clients and have special components are not displayed.

The following components are not displayed:

LdapServer, KrbServer, DBService, Hue, MapReduce, and Flume

----End

9.3.1.8.2 Batch Upgrading Clients

Scenario

The client package downloaded from FusionInsight Manager contains the client batch upgrade tool. When multiple clients need to be upgraded after the cluster upgrade or scale-out, you can use this tool to upgrade the clients in batches with a few clicks. In addition, the tool provides the lightweight function of batch updating the **/etc/hosts** file on the nodes where the clients are located.

Procedure

Prepare for the client upgrade.

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster**, click the name of the desired cluster, click **More**, and select **Download Client** to download the complete client to the specified directory on the server.

For details, see **Downloading the Client**.

Decompress the downloaded client package and find the **batch_upgrade** directory, for example, **/tmp/FusionInsight-Client/FusionInsight_Cluster_1_Services_ClientConfig/batch_upgrade**.

Step 3 Choose **Cluster**, click the name of the desired cluster, and choose **Client Management**. On the **Client Management** page, click **Export All** to export all client information to the local PC.

Step 4 Decompress the exported client information and upload the **client-info.cfg** file to the **batch_upgrade** directory.

Step 5 Supplement the password in the **client-info.cfg** file by referring to **Reference Information**.

Upgrade clients in batches.

Step 6 Run the **sh client_batch_upgrade.sh -u -f /tmp/FusionInsight-Client/FusionInsight_Cluster_1_Services_Client.tar -g /tmp/FusionInsight-Client/FusionInsight_Cluster_1_Services_ClientConfig/batch_upgrade/client-info.cfg** command to perform the upgrade.

NOTICE

You are advised to delete the **client-info.cfg** file as soon as possible after the upgrade because the password has been configured.

Step 7 After the upgrade is complete, verify the upgrade result by running the **sh client_batch_upgrade.sh -c** command.

Step 8 If the client is faulty, run the **sh client_batch_upgrade.sh -s** command to roll back the client.

NOTE

- The client batch upgrade tool moves the original client to the backup directory, and then uses the client package specified by the **-f** parameter to install the client. Therefore, if the original client contains customized content, manually save the customized content from the backup directory or move the customized content to the client directory after the upgrade before running the **-c** command. The backup path on the client is *{Original client path}*-**backup**.
- The **-u** command is the prerequisite for the **-c** and **-s** commands. You can run the **-c** command to commit the upgrade or the **-s** command to perform a rollback only after the **-u** command is executed to perform an upgrade.
- You can run the **-u** command multiple times to upgrade only the clients that fail to be upgraded.
- The client batch upgrade tool also supports the clients of earlier versions.
- When upgrading a client installed by a non-root user, ensure that the user has the read and write permissions on the directory where the client is located and the parent directory on the target node. Otherwise, the upgrade will fail.
- The client package specified by the **-f** parameter must be a full client package. The client packages of a single component or some components cannot be used as the input.

----End

Reference Information

Before upgrading clients in batches, you need to manually configure the user password for remotely logging in to the client node.

Run the **vi client-info.cfg** command to add a user password.

Example:

```
clientIp,clientPath,user,password
10.10.10.100,/home/omm/client /home/omm/client2,omm,Password
```

The fields in the configuration file are as follows:

- **clientIp**: indicates the IP address of the node where the client is located.
- **clientPath**: indicates the client installation path. Multiple paths are separated by spaces. Note that the path cannot end with a slash (/).
- **user**: indicates the username of the node.
- **password**: indicates the user password of the node.

NOTE

If the execution fails, view the **node.log** file in the **work_space/log_***XXX* directory.

9.3.1.8.3 Updating the hosts File in Batches

Scenario

The client package downloaded from FusionInsight Manager contains the client batch upgrade tool. This tool provides the function of upgrading clients in batches and the lightweight function of batch updating the **/etc/hosts** file on the node where the client is located.

Prerequisites

You have made preparations for the upgrade. For details, see "Prepare for the client upgrade." in **Batch Upgrading Clients**.

Updating the hosts File in Batches

Step 1 Check whether the user configured for the node where the **/etc/hosts** file needs to be updated is **root**.

- If yes, go to **Step 2**.
- If no, change the user to **root** and go to **Step 2**.

Step 2 Run the **sh client_batch_upgrade.sh -r -f /tmp/FusionInsight-Client/FusionInsight_Cluster_1_Services_Client.tar -g /tmp/FusionInsight-Client/FusionInsight_Cluster_1_Services_ClientConfig/batch_upgrade/client-info.cfg** command to batch update the **/etc/hosts** file on the nodes where the client resides.

NOTE

- When you batch update the **/etc/hosts** file, the entered client package can be a complete client package or a client package that contains only configuration files (recommended).
- The user configured for the host where the **/etc/hosts** file needs to be updated must be **root**. Otherwise, the update fails.

----End

9.3.2 Managing a Service

9.3.2.1 Overview

Dashboard

Log in to FusionInsight Manager. Choose **Cluster**, click the name of the desired cluster, and choose **Services**. The service management page is displayed, including the functional area and service list.

Functional Area

In the functional area of the service management page, you can select a view type and filter and search for services by service type. You can use the advanced search to select required services based on the running status and configuration status.

Service List

The service list on the service management page contains all installed services in the cluster. If the tile view is selected, the services will be displayed in pane style. If you select the list view, the services will be displayed in a table.

NOTE

In this section, the **Tile View** is used by default.

The service list displays the running status, configuration status, role type, and number of instances of each service. On this page, you can perform some service maintenance tasks, such as starting, stopping, and restarting services.

Table 9-5 Service running status

Status	Description
Normal	Indicates that the service is running properly.
Faulty	Indicates that the service cannot run properly.
Partially Healthy	Indicates that some enhanced functions of the service are abnormal.
Not started	Indicates that the service is stopped.
Unknown	Indicates that the initial status of the service cannot be detected.
Starting	Indicates that the service is being started.
Stopping	Indicates that the service is being stopped.
Failed to start	Indicates that the service fails to be started.
Failed to stop	Indicates that the service fails to be stopped.

NOTE

- If the running status of a service is **Faulty**, an alarm is generated. Rectify the fault based on the alarm information.
- HBase, Hive, Spark, and Loader may be in the **Subhealthy** state.
  - If Yarn is installed but is abnormal, HBase is in the **Subhealthy** state. If the multi-instance function is enabled, all installed HBase service instances are in the **Subhealthy** state.
  - If HBase is installed but is abnormal, Hive, Spark, and Loader are in the **Subhealthy** state.
  - If any HBase instance is installed but is abnormal after the multi-instance function is enabled, Loader is in the **Subhealthy** state.
  - If an HBase instance is installed but is abnormal after the multi-instance function is enabled, the Hive and Spark instances that map to the HBase instance are in the **Subhealthy** state. That is, if HBase 2 is installed but is abnormal, Hive 2 and Spark2 are in the **Subhealthy** state.

Table 9-6 Service configuration status

Status	Description
Synchronized	Indicates that all service parameter settings have taken effect in the cluster.

Status	Description
Expired	Indicates that the latest configuration is not synchronized and does not take effect after the service parameters are modified. You need to synchronize the configurations and restart the services. You can click ? next to **Configuration Status** to view expired configuration items.
Failed	Indicates that a communication or read/write exception occurs during the parameter configuration synchronization. Use **Synchronize Configuration** to rectify the fault.
Synchronizing	Indicates that the service parameter configuration is being synchronized.
Unknown	Indicates that the initial status of the service cannot be detected.

You can click a service in the service list to perform simple maintenance and management operations on the service, as described in **Table 9-7**.

Table 9-7 Basic maintenance and management

Menu Item on the UI	Description
Start Service	Start a specified service in the cluster.
Stop Service	Stop a specified service in the cluster.
Restart Service	Restart a specified service in the cluster. **NOTE** If a service is restarted, other services that depend on this service will be unavailable. Therefore, select **Restart upper-layer services**. Determine whether to perform this operation based on the displayed service list. Services are restarted one by one due to their dependency. **Table 9-8** describes the restart duration of a single service.
Service Rolling Restart	Restart a specified service in the cluster without interrupting services. For details about the parameter settings, see **Table 9-4**.
Synchronize Configuration	● Enable new configuration parameters for a specified service in the cluster. ● Distribute new configuration parameters for services whose **Configuration Status** is **Expired**. **NOTE** After some services are synchronized, restart the services for the settings to take effect.

Table 9-8 Restart duration

Service	Restart Duration	Startup Duration	Remarks
ClickHouse	4 min	ClickHouseServer: 2 min ClickHouseBalancer: 2 min	-
HDFS	10min+x	NameNode: 4 min + x DataNode: 2 min JournalNode: 2 min Zkfc: 2 min	*x* indicates the NameNode metadata loading duration. It takes about 2 minutes to load 10,000,000 files. For example, *x* is 10 minutes for 50 million files. The startup duration fluctuates with reporting of DataNode data blocks.
Yarn	5 min + x	ResourceManager: 3 min + x NodeManager: 2 min	*x* indicates the time required for restoring ResourceManager reserved tasks. It takes about 1 minute to restore 10,000 reserved tasks.
MapReduce	2 min + x	JobHistoryServer: 2 min + x	*x* indicates the scanning duration of historical tasks. It takes about 2.5 minutes to scan 100,000 tasks.
ZooKeeper	2 min + x	quorumpeer: 2 min + x	*x* indicates the duration for loading znodes. It takes about 1 minute to load 1 million znodes.
Hive	3.5 min	HiveServer: 3 min MetaStore: 1 min 30s WebHcat: 1 min Hive service: 3 min	-
Spark2x	5 min	JobHistory2x: 5 min SparkResource2x: 5 min JDBCServer2x: 5 min	-
Flink	4 min	FlinkResource: 1 min FlinkServer: 3 min	-

Service	Restart Duration	Startup Duration	Remarks
Kafka	2 min + x	Broker: 1 min + x	*x* indicates the data restoration duration. It takes about 2 minutes to start 20,000 partitions for a single instance.
Storm	6 min	Nimbus: 3 min UI: 1 min Supervisor: 1 min Logviewer: 1 min	-
Flume	3 min	Flume: 2 min MonitorServer: 1 min	-

9.3.2.2 Other Service Management Operations

9.3.2.2.1 Service Details Page

Overview

Log in to FusionInsight Manager and choose **Cluster** > *Name of the desired cluster* > **Services**. In the service list, click the specified service name to go to the service details page, including the **Dashboard**, **Instance**, **Instance Groups** and **Configurations** tab pages as well as function areas. For some services, the custom management tool page can be displayed.For details about the supported management tools, see **Table 9-9**..

Table 9-9 Customized management tools

Tool	Service	Description
Flume configuration tool	Flume	Configures collection parameters for the Flume server and client.
Flume client management tool	Flume	Views the monitoring information about the Flume client.
Kafka topic monitoring tool	Kafka	Monitors and manages Kafka topics.

The **Dashboard** page is the default page, which contains the basic information, role list, dependency table, and monitoring chart, and more. You can manage services in the upper right corner. For details about basic service management, such as starting, stopping, rolling restart, and synchronization configuration, see **Table 9-7**. For details about other service management operations, see **Table 9-10**.

Table 9-10 Service management operations

Navigation Path	Description
More > **Health Check**	Performs a health check for the current service. The health check items include the health status of each check object, related alarms, and user-defined monitoring indicators. The check result is not the same as the values of **Running Status** displayed on the GUI. To export the result of the health check, click **Export Report** in the upper left corner of the checklist. If you find any problem, click **View Help**.
More > **Download Client**	Download the default client that contains only specific services and perform management operations, run services, or perform secondary development on the client. For details, see **Downloading the Client**.
More > **Change Service Name**	Changes the name of the current service.
More > **Perform** *XX* **Switchover**	For details, see **Performing Active/Standby Switchover of a Role Instance**.
More > **Enter/Exit Maintenance Mode**	Configures a service to enter/exit the maintenance mode.
Configurations > **Import/Export**	In the scenario where services are migrated to a new cluster or the same services are deployed again, you can import or export all configuration data of a specific service to quickly copy the configuration results.

Basic Information Area

The basic information area on the **Dashboard** tab page contains the basic status data of the service, including the running status, configuration details, version, and key information of the service. If the service supports the open-source web UIs, you can access the open-source web UIs by clicking the links in the basic information area.

NOTE

In the current version, user **admin** does not have the permission to access all the service functions provided on the open source web UI. Create a component service administrator to access the WebUI address.

Role List

The role list on the **Dashboard** tab page contains all roles of the service. The role list displays the running status and the number of instances of each role.

Dependency

The dependency relationship table on the **Dashboard** tab page displays the services on which the current service depends and other services that depend on the service.

Historical Records of Alarms and Events

The alarm and event history area displays the key alarms and events reported by the current service. Up to 20 historical records are displayed.

Chart

The chart area is displayed on the right of the **Dashboard** tab page and contains the key monitoring indicator report of the service. You can customize the monitoring report that is displayed in the chart area, view the description of the monitoring metrics, or export the monitoring data. For a customized resource contribution chart, you can zoom in on the chart and switch between the trend chart and distribution chart.

NOTE

Some services in the cluster provide service-level resource monitoring items. For details, see **Resource Monitoring**.

9.3.2.2.2 Performing Active/Standby Switchover of a Role Instance

Scenario

Some service roles are deployed in active/standby mode. If the active instance needs to be maintained and cannot provide services, or other maintenance is required, you can manually trigger an active/standby switchover.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 3 Click the specified service name on the service management page.

Step 4 On the service details page, expand the **More** drop-down list and select **Perform *Role Instance* Switchover**.

Step 5 In the displayed dialog box, enter the password of the current login user and click **OK**.

Step 6 In the displayed dialog box, click **OK** to perform active/standby switchover for the role instance.

NOTE

- The Manager component package only supports the active/standby switchover of DBService role instances.
- The HD component package supports the active/standby switchover of the following service role instances: HDFS, YARN, Storm, HBase, and MapReduce.
- When an active/standby switchover is performed for a NameNode on HDFS, a NameService must be set.
- The Porter component package only supports the active/standby switchover of Loader role instances.
- This function cannot be used for other role instances.

----End

9.3.2.2.3 Resource Monitoring

Log in to FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services**, and click **Resource**. The resource monitoring page is displayed.

Some services in the cluster provide service-level resource monitoring metrics. By default, the monitoring data of the latest 12 hours is displayed. You can click ✎ to customize a time range. Time range options are **12h**, **1d**, **1w**, and **1m**. You can click ↗ to export the corresponding report information. If a monitoring item has no data, the report cannot be exported. **Table 9-11** lists the services and monitoring items that support resource monitoring.

Table 9-11 Service resource monitoring

Service	Metrics	Description
HDFS	Resource Usage (by Tenant)	● Collects statistics on HDFS resource usage by tenant. ● Views the metrics **Capacity** or **Number of File Objects**.
	Resource Usage (by User)	● Collects statistics on HDFS resource usage by user. ● Views the metrics **Used Capacity** or **Number of File Objects**.

Service	Metrics	Description
	Resource Usage (by Directory)	● Collects statistics on HDFS resource usage by directory. ● Views the metrics **Used Capacity** or **Number of File Objects**. ● You can click ▤ to configure space monitoring. Alternatively, you can specify an HDFS file system directory for monitoring.
	Resource Usage (by Replica)	● Collects statistics on HDFS resource usages by replica count. ● Views the metrics **Used Capacity** or **File Count**.
	Resource Usage (by File Size)	● Collects statistics on HDFS resource usages by file size. ● Views the metrics **Used Capacity** or **File Count**.
	Recycle Bin (by User)	● Collects statistics on the usage of the HDFS recycle bin by user. ● Views the metrics **Recycle Bin Capacity** or **Number of File Objects**.
	Operation Count	● Collects the number of operations in HDFS.
	Automatic Balancer	● Collects statistics on the execution speed of HDFS automatic balancer and the total capacity of the current balancer migration.
	NameNode RPC Open Connections (by User)	● Displays the number of connections of each user in the Client RPC requests connected to NameNodes.
	Slow DataNodes	Displays DataNode that transmits or processes data slowly in the cluster.
	Slow Disks	Displays the disk that processes data slowly on the DataNode in the cluster.
HBase	Operation Requests in Tables	Displays the number of PUT, DELETE, GET, SCAN, INCREMENT, and APPEND operation requests in all tables on all RegionServers.

Service	Metrics	Description
	Operation Requests on RegionServers	Displays the number of PUT, DELETE, GET, SCAN, INCREMENT, and APPEND operation requests and number of all operation requests in RegionServer.
	Operation Requests for Service	Displays the number of PUT, DELETE, GET, SCAN, INCREMENT, and APPEND operation requests in all regions on RegionServers.
	HFiles on RegionServers	Displays the number of HFiles in all RegionServers.
Hive	HiveServer2-Background-Pool Threads (by IP)	Displays the number of HiveServer2-Background-Pool threads of top users. These threads are measured and displayed in a measurement period.
	HiveServer2-Handler-Pool Threads (by IP)	Displays the number of HiveServer2-Handler-Pools of top users collected and displayed in a period.
	Used MetaStore Number (by IP)	Collects statistics on and displays the MetaStore usage of top users in a period.
	Number of Hive jobs	Displays the number of user-related jobs collected by Hive in a period.
	Number of Files Accessed in the Split Phase	Displays the number of files accessed by the underlying file storage system (HDFS by default) in the Split phase in a period.
	Hive Basic Operation Time	Collects time for creating a directory (mkdirTime), creating a file (touchTime), writing a file (writeFileTime), renaming a file (renameTime), moving a file (moveTime), deleting a file (deleteFileTime), and deleting a directory (deleteCatalogTime) in a period of time.
	Table Partitions	Displays the number of partitions in all Hive tables, which is displayed in the following format: *database # table name*, *number of table partitions*.

Service	Metrics	Description
	HQL Map Count	Collects statistics on HQL statements executed in a period and the number of Map statements invoked during the execution. The displayed information includes users, HQL statements, and the number of Map statements.
	HQL Access Statistics	Displays the number of HQL access times in a period.
Kafka	Kafka Disk Usage Distribution	Displays the disk usage distribution statistics of the Kafka cluster.
Spark2x	HQL Access Statistics	Collects HQL access statistics in a period, including the username, HQL statement, and HQL statement execution times.
Yarn	Used resources (by task)	● Displays the number of CPU cores and memory used by a task. ● Views the metrics **By memory** or **By CPU**.
	Resource usage (by tenant)	● Displays the number of CPU cores and memory used by a tenant. ● Views the metrics **By memory** or **By CPU**.
	Resource usage ratio (by tenant)	● Displays the ratio of the number of CPU cores to the memory used by a tenant. ● Views the metrics **By memory** or **By CPU**.
	Task Duration Ranking	Displays Yarn tasks sorted by time consumption.
	ResourceManager RPC Open Connections (by User)	Displays the number of client RPC connections to ResourceManager by user.
	Operation Count	Collects statistics on the number and proportion of operations corresponding to each Yarn operation type.
	Ranking of Tasks in a Queue by Resource Usage	● Displays the resources consumed by the tasks running in a queue after the queue (tenant) is selected on the GUI. ● Views the metrics **By memory** or **By CPU**.

Service	Metrics	Description
	Ranking of Users in a Queue by Resource Usage	● Displays the resources consumed by the users who are running tasks in the queue after a queue (tenant) is selected on the GUI. ● Views the metrics **By memory** or **By CPU**.
ZooKeeper	Used Resources (By Second-Level Znode)	● Displays the ZooKeeper level-2 znode resource status. ● Views the metrics **By Znode quantity** or **By capacity**.
	Number of Connections (by Client IP Address)	Displays the ZooKeeper client connection resource status.

9.3.2.2.4 Collecting Stack Information

Scenario

To meet actual service requirements, the cluster administrator can collect stack information about a specified role or instance on FusionInsight Manager, save the information to a local directory, and download the information. The following information can be collected:

1. jstack information.
2. jmap -histo information.
3. jmap -dump information.
4. Thr jstack and jmap-histo information can be collected continuously for comparison.

Procedure

Collecting stack information

Step 1 Log in to FusionInsight Manager.

Step 2 Click **Cluster**, click the name of the desired cluster, click **Services**, and click the target service.

Step 3 On the displayed page, Choose **More** > **Collect Stack Information**.

NOTE

- To collect stack information of multiple instances, go to the instance list, select the desired instances in the instance list and choose **More** > **Collect Stack Information**.
- To collect stack information of a single instance, click the desired instance and choose **More** > **Collect Stack Information**.

Step 4 In the displayed dialog box, select the desired role and content, configure advanced options (retain the default settings if there is no special requirement), and click **OK**.

Step 5 After the collection is successful, click **Download**.

Downloading Stack Information

Step 6 Click **Cluster**, click the name of the desired cluster, click **Services**, and click the target service. Choose **More** > **Download Stack Information** in the upper right corner.

Step 7 Select the desired role and content and click **Download** to download the stack information to the local PC.

Clearing stack information

Step 8 Click **Cluster**, click the name of the desired cluster, click **Services**, and click the target service.

Step 9 Choose **More** > **Clear Stack Information** in the upper right corner.

Step 10 Select the desired role and content and configure **File Directory**. Click **OK**.

----End

9.3.2.2.5 Switching Ranger Authentication

Scenario

By default, the Ranger service is installed and Ranger authentication is enabled for a newly installed cluster in security mode. You can set fine-grained security access policies for accessing component resources through the permission plug-in of the component. If Ranger authentication is not required, the cluster administrator can manually disable Ranger authentication on the service page. After Ranger authentication is disabled, the system continues to perform permission control based on the role model of FusionInsight Manager when accessing component resources.

In a cluster upgraded from an earlier version, Ranger authentication is not used by default when users access component resources. The cluster administrator can manually enable Ranger authentication after installing the Ranger service.

NOTE

- In a cluster in security mode, the following components support Ranger authentication: HDFS, YARN, Kafka, Hive, HBase, Storm, Impala, and Spark2x.
- In a cluster in non-security mode, Ranger supports permission control on component resources based on OS users. The following components support Ranger authentication: HBase, HDFS, Hive, Spark2x, and YARN.
- After Ranger authentication is enabled, all authentication of the component will be managed by Ranger. The permissions set by the original authentication plug-in will become invalid (The ACL rules of HDFS and YARN components still take effect). Exercise caution when performing this operation. You are advised to deploy permissions on Ranger in advance.
- After Ranger authentication is disabled, all authentication of the component will be managed by the permission plug-in of the component. The permission set on Ranger will become invalid. Exercise caution when performing this operation. You are advised to deploy permissions on Manager in advance.

Enabling Ranger Authentication

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > **Services**.

Step 3 Click the specified service name on the service management page.

Step 4 On the service details page, expand the **More** drop-down list and select **Enable Ranger**.

Step 5 In the displayed dialog box, enter the password of the current login user and click **OK**.

Step 6 In the service list, restart the service whose configuration has expired.

----End

Disabling Ranger Authentication

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > **Services**.

Step 3 Click the specified service name on the service management page.

Step 4 On the service details page, expand the **More** drop-down list and select **Disable Ranger**.

Step 5 Enter the password of the current login user and click **OK**. In the displayed dialog box, click **OK**.

Step 6 In the service list, restart the service whose configuration has expired.

----End

9.3.2.3 Service Configuration

9.3.2.3.1 Modifying Service Configuration Parameters

Scenario

To meet actual service requirements, cluster administrators can quickly view and modify default service configurations on FusionInsight Manager. Configure parameters based on the information provided in the configuration description.

NOTE

The parameters of DBService cannot be modified when only one DBService role instance exists in the cluster.

Impact on the System

- After configuring properties of a service, you need to restart the service if the service status is **Expired**. The service is unavailable during the restart.
- After the service configuration parameters are modified and then take effect after restart, you need to download and install the client again or download

the configuration file to update the client. For example, you can modify configuration parameters of the following services: HBase, HDFS, Hive, Spark, YARN, and MapReduce.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 3 Click the specified service name on the service management page.

Step 4 Click **Configuration**.

The **Basic Configuration** page is displayed by default. To modify more parameters, click the **All Configurations** tab. The navigation tree displays all configuration parameters of the service. The level-1 nodes in the navigation tree are service names or role names. The parameter category is displayed after the level-1 node is expanded.

Step 5 In the navigation tree, select the specified parameter category and change the parameter values on the right.

NOTE

Select a port parameter value from the value range on the right. Ensure that all parameter values in the same service are within the value range and are unique. Otherwise, the service fails to be started.

If you are not sure about the location of a parameter, you can enter the parameter name in search box in the upper right corner. The system searches for the parameter in real time and displays the result.

Step 6 Click **Save**. In the confirmation dialog box, click **OK**.

Wait until the message "Operation succeeded." is displayed. Click **Finish**.

The configuration is modified.

NOTE

- To update the queue configuration of the YARN service without restarting service, choose **More** > **Refresh Queue** to update the queue for the configuration to take effect.
- During configuration of the **flume.config.file** parameter, you can upload and download files. After a configuration file is uploaded, the old file will be overwritten. If the configuration is not saved and the service is restarted, the configuration does not take effect. Save the configuration in time.
- If you need to restart the service for the configuration to take effect after modifying service configuration parameters, choose **More** > **Restart Service** in the upper right corner of the service page.

----End

9.3.2.3.2 Modifying Custom Configuration Parameters of a Service

Scenario

All open source parameters can be configured for all MRS cluster components. Parameters used in some key application scenarios can be modified on

FusionInsight Manager, and some parameters of open source features may not be configured for some component clients. To modify the component parameters that are not directly supported by Manager, cluster administrators can add new parameters for components using the configuration customization function on Manager. Newly added parameters are saved in component configuration files and take effect after restart.

Impact on the System

- After configuring properties of a service, you need to restart the service if the service status is **Expired**. The service is unavailable during the restart.
- After the service configuration parameters are modified and then take effect after restart, you need to download and install the client again or download the configuration file to update the client.

Prerequisites

Cluster administrators have fully understood the meanings of the parameters to be added, configuration files to take effect, and the impact on components.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 3 Click the specified service name on the service management page.

Step 4 Click **Configuration** and click **All Configurations**.

Step 5 In the navigation tree on the left, locate a level-1 node and select **Customization**. The system displays the customized parameters of the current component.

The configuration files that save the newly added custom parameters are displayed in the **Parameter File** column. Different configuration files may have same open source parameters. After the parameters in different files are set to different values, the configuration takes effect depends on the loading sequence of the configuration files by components. You can customize parameters for services and roles as required. Adding customized parameters for a single role instance is not supported.

Step 6 Locate the row where a specified parameter resides, enter the parameter name supported by the component in the **Name** column and enter the parameter value in the **Value** column.

You can click **+** or **-** to add or delete a customized parameter.

Step 7 Click **Save**. In the displayed **Save Configuration** dialog box, confirm the modification and click **OK**. After the system displays "Operation succeeded", click **Finish**. The configuration is saved successfully.

Restart the expired service or instance for the configuration to take effect.

----End

Task Example (Configuring Customized Hive Parameters)

Hive depends on HDFS. By default, Hive accesses the HDFS client. The configuration parameters that have taken effect are controlled by HDFS. For example, the HDFS parameter **ipc.client.rpc.timeout** affects the RPC timeout interval for all clients to connect to the HDFS server. Cluster administrators can modify the timeout interval for Hive to connect to HDFS by configuring custom parameters. After this parameter is added to the **core-site.xml** file of Hive, this parameter can be identified by the Hive service and its configuration overwrites the parameter configuration in HDFS.

Step 1 On FusionInsight Manager, click **Cluster**, click the name of the desired cluster, and click **Services**.

Step 2 On the displayed page, click **Configuration** and click **All Configurations**.

Step 3 In the navigation tree on the left, select **Customization** for the Hive service. The system displays the custom service parameters supported by Hive.

Step 4 In **core-site.xml**, locate the row that contains the **core.site.customized.configs** parameter, enter **ipc.client.rpc.timeout** in the **Name** column, and enter a new value in the **Value** column, for example, 150000. The unit is ms.

Step 5 Click **Save**. In the displayed **Save Configuration** dialog box, confirm the modification and click **OK**. Wait until the message "Operation succeeded" is displayed, and click **Finish**.

The configuration is saved successfully.

After the configuration is saved, restart the expired service or instance for the configuration to take effect.

----End

9.3.3 Instance Management

9.3.3.1 Overview

Overview

Log in to FusionInsight Manager, click **Cluster**, click the name of the desired cluster, and choose **Service** > **KrbServer**. On the displayed page, click **Instance**. The displayed instance management page contains the function area and role instance list.

Functional Area

After selecting the instances to be operated in the function area, you can maintain and manage the role instances, such as starting or stopping the instances. **Table 9-12** shows the main operations.

Table 9-12 Instance maintenance and management

UI Portal	Description
Start Instance	Start a specified instance in the cluster. You can start a role instance in the **Not Started**, **Stop Failed**, or **Startup Failed** state to use the role instance.
More > Stop Instance	Stop a specified instance in the cluster. You can stop a role instance that is no longer used or is abnormal.
More > Restart Instance	Restart a specified instance in the cluster. You can restart an abnormal role instance to restore it.
More > Instance Rolling Restart	Restart a specified instance in the cluster without interrupting services. For details about the parameter settings, see **Performing a Rolling Restart of a Cluster**.
More > Decommission/ Recommission	Recommission or decommission a specified instance in the cluster to change the service availability status of the service. For details, see **Decommissioning and Recommissioning an Instance**. NOTE Only the role DataNode in HDFS, role NodeManager in Yarn, and role RegionServer in HBase support the recommissioning and decommissioning functions.
Desired instance > **More** > **Synchronize Configuration**	If the **Configuration Status** of a role instance is **Expired**, the role instance has not been restarted after the configuration is modified, and the new configuration is saved only on FusionInsight Manager. In this case, use this function to deliver the new configuration to the specified instance. NOTE ● After synchronizing the role instance configuration, you need to restart the role instance whose configuration has expired. The role instance is unavailable during the restart. ● After the synchronization is complete, restart the instance for the configuration to take effect.
Desired instance > **Instance Configurations**	For details, see **Managing Instance Configurations**.

You can filter instances based on the role they belong to or their running status in the function area.

NOTE

Click **Advanced Search** to search for specified instances by specifying other filter criteria, such as **Host Name**, **Management IP Address**, **Business IP Address**, or **Instance Groups**.

Role Instance List

The role instance list contains the instances of all roles in the cluster. The list displays the running status, configuration status, hosts, and related IP addresses of each instance.

Table 9-13 Instance running status

Status	Description
Normal	Indicates that the instance is running properly.
Faulty	Indicates that the instance cannot run properly.
Decommissioned	Indicates that the instance is out of service.
Not started	Indicates that the instance is stopped.
Unknown	Indicates that the initial status of the instance cannot be detected.
Starting	Indicates that the instance is being started.
Stopping	Indicates that the instance is being stopped.
Restoring	Indicates that an exception may occur in the instance and the instance is being automatically rectified.
Decommissioning	Indicates that the instance is being decommissioned.
Recommissioning	Indicates that the instance is being recommissioned.
Failed to start	Indicates that the service fails to be started.
Failed to stop	Indicates that the service fails to be stopped.

Instance Details

You can click an instance name to go to the instance details page and view the basic information, configuration file, instance logs, and monitoring metric reports of the instance.

9.3.3.2 Decommissioning and Recommissioning an Instance

Scenario

Some role instances provide services for external services in distributed and parallel mode. Services independently store information about whether each instance can be used. Therefore, you need to use FusionInsight Manager to recommission or decommission these instances to change the instance running status.

Some instances do not support the recommissioning and decommissioning functions.

NOTE

The following roles support decommissioning and recommissioning: HDFS DataNode, YARN NodeManager, and HBase RegionServer.

- If the number of the DataNodes is less than or equal to that of HDFS replicas, decommissioning cannot be performed. If the number of HDFS replicas is three and the number of DataNodes is less than four in the system, decommissioning cannot be performed. In this case, an error will be reported and force FusionInsight Manager to exit the decommissioning 30 minutes after FusionInsight Manager attempts to perform the decommissioning.
- During MapReduce task execution, files with 10 replicas are generated. Therefore, if the number of DataNode instances is less than 10, decommissioning cannot be performed.
- If the number of DataNode racks (the number of racks is determined by the number of racks configured for each DataNode) is greater than 1 before the decommissioning, and after some DataNodes are decommissioned, that of the remaining DataNodes changes to 1, the decommissioning will fail. Therefore, before decommissioning DataNode instances, you need to evaluate the impact of decommissioning on the number of racks to adjust the DataNodes to be decommissioned.
- If multiple DataNodes are decommissioned at the same time, and each of them stores a large volume of data, the DataNodes may fail to be decommissioned due to timeout. To avoid this problem, it is recommended that one DataNode be decommissioned each time and multiple decommissioning operations be performed.

Procedure

Step 1 Perform the following steps to perform a health check for the DataNodes before decommissioning:

1. Log in to the client installation node as a client user and switch to the client installation directory.
2. For a security cluster, use user **hdfs** for permission authentication.
```
source bigdata_env              #Configure client environment variables.
kinit hdfs                  #Configure kinit authentication.
Password for hdfs@HADOOP.COM:    #Enter the login password of user hdfs.
```
3. Run the **hdfs fsck / -list-corruptfileblocks** command, and check the returned result.
   - If "has 0 CORRUPT files" is displayed, go to **Step 2**.
   - If the result does not contain "has 0 CORRUPT files" and the name of the damaged file is returned, go to **Step 1.4**.
4. Run the **hdfs dfs -rm** *Name of the damaged file* command to delete the damaged file.

   NOTE

   Deleting a file or folder is a high-risk operation. Ensure that the file or folder is no longer required before performing this operation.

Step 2 Log in to FusionInsight Manager.

Step 3 Choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 4 Click the specified service name on the service management page. On the displayed page, click the **Instance** tab.

Step 5 Select the specified role instance to be decommissioned.

Step 6 Select **Decommission** or **Recommission** from the **More** drop-down list.

In the displayed dialog box, enter the password of the current login user and click **OK**.

Select **I confirm to decommission these instances and accept the consequence of service performance deterioration** and click **OK** to perform the corresponding operation.

NOTE

During the instance decommissioning, if the service corresponding to the instance is restarted in the cluster using another browser, FusionInsight Manager displays a message indicating that the instance decommissioning is stopped, but the operating status of the instance is displayed as **Started**. In this case, the instance has been decommissioned on the background. You need to decommission the instance again to synchronize the operating status.

----End

9.3.3.3 Managing Instance Configurations

Scenario

Configuration parameters of each role instance can be modified. In the scenario where instances are migrated to a new cluster or the service is redeployed, the cluster administrator can import or export all configuration data of a service on FusionInsight Manager to quickly copy configuration results.

FusionInsight Manager can manage configuration parameters of a single role instance. Modifying configuration parameters and importing or exporting instance configurations do not affect other instances.

Impact on the System

After modifying the configuration of a role instance, you need to restart the instance if the instance status is **Expired**. The role instance is unavailable during restart.

Modifying Instance Configuration

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 3 On the page that is displayed, click the **Instance** tab.

Step 4 Click the specified instance and select **Instance Configuration**.

By default, **Basic Configuration** is displayed. To modify more parameters, click **All Configurations**. All parameter categories supported by the instance are displayed on the **All Configurations** tab page.

Step 5 In the navigation tree, select the specified parameter category and change the parameter values on the right.

If you are not sure about the location of a parameter, you can enter the parameter name in search box in the upper right corner. The system searches for the parameter in real time and displays the result.

Step 6 Click **Save**. In the confirmation dialog box, click **OK**.

Wait until the message "Operation succeeded." is displayed. Click **Finish**.

The configuration is modified.

NOTE

After the configuration parameters of a role instance are modified, you need to restart the instance if the instance status is **Expired**. You can select the expired instance on the **Instances** page and choose **More** > **Restart Instance**.

----End

Exporting/Importing Instance Configuration

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 3 Click the specified service name on the service management page. On the displayed page, click the **Instance** tab.

Step 4 Click the specified instance and select **Instance Configurations**.

Step 5 Click **Export** to export the configuration parameter file to the local host.

Step 6 On the **Instance Configurations** page, click **Import**, select the configuration parameter file of the instance, and import the file.

----End

9.3.3.4 Viewing the Instance Configuration File

Scenario

FusionInsight Manager allows O&M personnel to view the content configuration files such as environment variables and role configurations of the instance node on the management page. If O&M personnel need to quickly check whether configuration items of the instance are incorrectly configured or when some hidden configuration items need to be viewed, the O&M personnel can directly view the configuration files on FusionInsight Manager. In this case, users quickly analyze configuration problems.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Service**.

Step 3 Click the specified service name on the service management page. On the displayed page, click the **Instance** tab.

Step 4 Click the name of the target instance. In the **Configuration File** area on the **Instance Status** page, the configuration file list of the instance is displayed.

Step 5 Click the name of the configuration file to be viewed to view the parameter values in the configuration file.

To obtain the configuration file, you can download the configuration file to the local PC.

NOTE

If a node in the cluster is faulty, the configuration file cannot be viewed. Rectify the fault before viewing the configuration file again.

----End

9.3.3.5 Instance Group

9.3.3.5.1 Managing Instance Groups

Scenario

Instance groups can be managed on FusionInsight Manager. That is, you can group multiple instances in the same role based on a specified principle, such as the nodes with the same hardware configuration. The modification on the configuration parameters of an instance group applies to all instances in the group.

In a large cluster, instance groups are used to improve the capability of managing instances in batches in the heterogeneous environment. After instances are grouped, the instances can be configured repeatedly to reduce redundant instance configuration items and improve system performance.

Creating an Instance Group

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 3 Click the specified service name on the service management page.

Step 4 On the displayed page, click the **Instance Groups** tab.

Click ⊕ and configure parameters as prompted.

Table 9-14 Instance group configuration parameters

Parameter	Description
The group name	Indicates the instance group name. The value can contain only letters, digits, underscores (_), hyphens (-), and spaces. It must start with a letter, digit, underscore (_), or hyphen (-) and cannot ends with a space. It can contain a maximum of 99 characters.

Parameter	Description
Role	Indicates the role to which an instance group belongs.
Copy From	Indicates that the parameter values of a specified instance group are copied to the parameters of a new group. If the value is null, the default values are used for the parameters of the new group.
Description	Indicates the instance group description. It can contain only letters, digits, commas (,), periods (.), underscores (_), spaces, and line breaks, and can contain a maximum of 200 characters.

NOTE

- Each instance must belong to only one instance group. When an instance is installed for the first time, it belongs to the instance group *Role name*-**DEFAULT** by default.
- You can delete unnecessary or unused instance groups. Before deleting an instance group, migrate all instances in the group to other instance groups, and then delete the instance group by referring to **Deleting an Instance Group**. The default instance group cannot be deleted.

Step 5 Click **OK**.

The instance group is created.

----End

Modifying Properties of an Instance Group

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 3 Click the specified service name on the service management page.

Step 4 Click the **Instance Groups** tab. On the **Instance Groups** tab page, locate the row that contains the target instance group.

Click ✎ and modify parameters as prompted.

Step 5 Click **OK** to save the modifications.

The default instance group cannot be modified.

----End

Deleting an Instance Group

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 3 Click the specified service name on the service management page.

Step 4 Click the **Instance Groups** tab. On the **Instance Groups** tab page, locate the row that contains the target instance group.

Step 5 Click 🗑.

Step 6 In the displayed dialog box, click **OK**.

The default instance group cannot be deleted.

----End

9.3.3.5.2 Viewing Information About an Instance Group

Scenario

The cluster administrator can view the instance group of a specified service on FusionInsight Manager.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 3 Click the specified service name on the service management page.

Step 4 On the displayed page, click the **Instance Groups** tab.

Step 5 In the navigation tree, select a role. On the **Basic** tab page, view all instances in the instance group.

NOTE

To move an instance from an instance group to another, perform the following operations:

1. Select the instance to be moved and click **Move**.
2. In the displayed dialog box, select an instance group to which the instance to be moved.

   During the migration, the configuration of the new instance group is automatically inherited. If the instance configuration is modified before the migration, the configuration of the instance prevails.
3. Click **OK**.

   Restart the expired service or instance for the configuration to take effect.

----End

9.3.3.5.3 Configuring Instantiation Group Parameters

Scenario

In a large cluster, users can configure parameters for multiple instances in batches by configuring the related instance groups on FusionInsight Manager, reducing redundant instance configuration items and improving system performance.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 3 Click the specified service name on the service management page.

Step 4 On the displayed page, click the **Instance Groups** tab.

Step 5 In the navigation tree, select the instance group name of a role, and switch to the **Configuration** tab page. Adjust parameters to be modified, and click **Save**. The configuration takes effect for all instances in the instance group.

----End

9.4 Hosts

9.4.1 Host Management Page

9.4.1.1 Viewing the Host List

Overview

Log in to FusionInsight Manager, click **Hosts**, and the host list is displayed on the host management page. You can view the host list and basic information of each host.

You can switch view types and set search criteria to filter and search for hosts.

Host View

You can click **Role View** to view the roles deployed on each host. If the role supports the active/standby mode, the role name is displayed in bold.

Host List

The host list on the host management page contains all hosts in the cluster, and O&M operations can be performed on these hosts.

On the host management page, you can filter hosts by node type or cluster. The rules for filtering host types are as follows:

- A management node is the node where OMS is deployed. Additionally, control roles and data roles may also be deployed on management nodes.
- A control node is the node where control roles are deployed. Additionally, data roles may also be deployed on control nodes.
- A Data Node is the node where only data roles are deployed.

If you select the **Host View**, the IP address, rack planning, AZ name, running status, cluster name, and hardware resource usage of each host are displayed.

Table 9-15 Host running status

Status	Description
Normal	Indicates that the host is in the normal state.
Faulty	Indicates that the host is abnormal.
Unknown	Indicates that the initial status of the host cannot be detected.
Isolated	Indicates that the host is isolated.
Suspended	Indicates that the host is stopped.

9.4.1.2 Viewing the Host Dashboard

Overview

Log in to FusionInsight Manager, click **Hosts**, and click a host name in the host list. The host details page contains the basic information area, disk status area, role list area, and monitoring chart.

Basic Information Area

The basic information area contains the key information about the host, such as the management IP address, service IP address, host type, rack, firewall, number of CPU cores, and OS.

Disk Status Area

The disk status area contains all disk partitions configured for the cluster on the host and the usage of each disk partition.

Instance List Area

The instance list area displays all role instances installed on the host and the status of each role instance. You can click the log file next to a role instance name to view the log file content of the instance online.

Alarm and Event History

The alarm and event history area displays the key alarms and events reported by the current host. The system can display a maximum of 20 historical records.

Chart

The monitoring chart area is displayed on the right of the host details page, and contains the key monitoring metrics of the host.

You can choose ▼ > **Customize** in the upper right corner to customize the monitoring reports to be displayed in the chart area. Select a time range and choose ▼ > **Export** to export detailed monitoring metric data within the specified time range.

You can click ⓘ next to the title of a monitoring indicator to open the description of the monitoring indicator.

Click the **Chart** tab of the host to view the full monitoring chart information about the host.

GPU Card Status Area

If the host is configured with GPU cards, the GPU card status area displays the model, location, and status of the GPU card installed on the host.

9.4.1.3 Checking Host Processes and Resources

Overview

Log in to FusionInsight Manager, click **Hosts**, and click the specified host name in the host list. On the host details page, click the **Process** and **Resource** tabs.

Host Process

On the **Process** tab page, the information about the role processes of the deployed service instances on the current host is displayed, including the process status, PID, and process running time. You can directly view the log files of each process online.

Host Resource

On the **Resource** tab page, the detailed resource usage of deployed service instances on the current host is displayed, including the CPU, memory, disk, and port usage.

9.4.2 Host Maintenance Operations

9.4.2.1 Starting and Stopping All Instances on a Host

Scenario

If a host is faulty, you may need to stop all the roles on the host and perform maintenance check on the host. After the host fault is rectified, start all roles running on the host to recover host services. You can start or stop all instances on a host on the host management page or host details page on FusionInsight Manager. The following describes how to perform such operations on the host management page.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Click **Hosts**.

Step 3 Select the check box of the target host.

Step 4 Select **Start All Instances** or **Stop All Instances** from the **More** drop-down list to start or stop all role instances.

----End

9.4.2.2 Performing a Host Health Check

Scenario

If the running status of a host is not **Normal**, you can perform health checks on the host to check whether some basic functions are abnormal. During routine O&M, you can perform host health checks to ensure that the configuration parameters and monitoring of each role instance on the host are normal and can run stably for a long time.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Click **Hosts**.

Step 3 Select the check box of the target host.

Step 4 Select **Health Check** from the **More** drop-down list to start the health check.

To export the result of the health check, click **Export Report** in the upper left corner. If any problem is detected, click **Help**.

----End

9.4.2.3 Configuring Racks for Hosts

Scenario

All hosts in a large cluster are usually deployed on multiple racks. Hosts on different racks communicate with each other through switches. The network bandwidth between different hosts on the same rack is much greater than that on different racks. In this case, plan the network topology based on the following requirements:

- To improve the communication speed, it is recommended that data be exchanged between hosts on the same rack.
- To improve the fault tolerance capability, distribute processes or data of distributed services on different hosts of multiple racks as dispersedly as possible.

Hadoop uses a file directory structure to represent hosts.

The HDFS cannot automatically determine the network topology of each DataNode in the cluster. You need to set the rack name to identify the rack where the host is located so that the NameNode can draw the network topology of the required DataNodes and back up data of the DataNodes to different racks.

Similarly, YARN needs to obtain rack information and allocate tasks to different NodeManagers as required.

If the cluster network topology changes, you need to reallocate racks for hosts on FusionInsight Manager so that related services can be automatically adjusted.

Impact on the System

If the name of the host rack is changed, storage policy for HDFS replicas, YARN task assignment, and storage location of Kafka partitions will be affected. After the modification, you need to restart the HDFS, YARN, and Kafka for the configuration to take effect.

Improper rack configuration will unbalance loads (including CPU, memory, disk, and network) among nodes in the cluster, which decreases the cluster reliability and stability. Therefore, before allocating racks, take all aspects into consideration and properly set racks.

Rack Allocation Policies

NOTE

Physical rack: indicates the real rack where the host resides.

Logical rack: indicates the rack name of the host on FusionInsight Manager.

Policy 1: Each logical rack has nearly the same number of hosts.

Policy 2: The name of the logical rack of the host must comply with that of the physical rack to which the host belongs.

Policy 3: If there are only few hosts on a physical rack, combine this physical rack and other physical racks with few hosts into a logical rack, which complies with policy 1. Hosts in two equipment rooms cannot be placed in one logical rack. Otherwise, performance problems may be caused.

Policy 4: If there are lots of hosts on a physical rack, divide these hosts into multiple logical racks, which complies with policy 1. Hosts with great differences should not be placed in the same logical rack. Otherwise, the cluster reliability will be decreased.

Policy 5: You are advised to set **default** or other values for logical racks on the first layer, and the values in the same cluster must be consistent.

Policy 6: The number of hosts in each rack cannot be less than 3.

Policy 7: A cluster can contain at most 50 logical racks. If there are too many logical racks in a cluster, the maintenance is difficult.

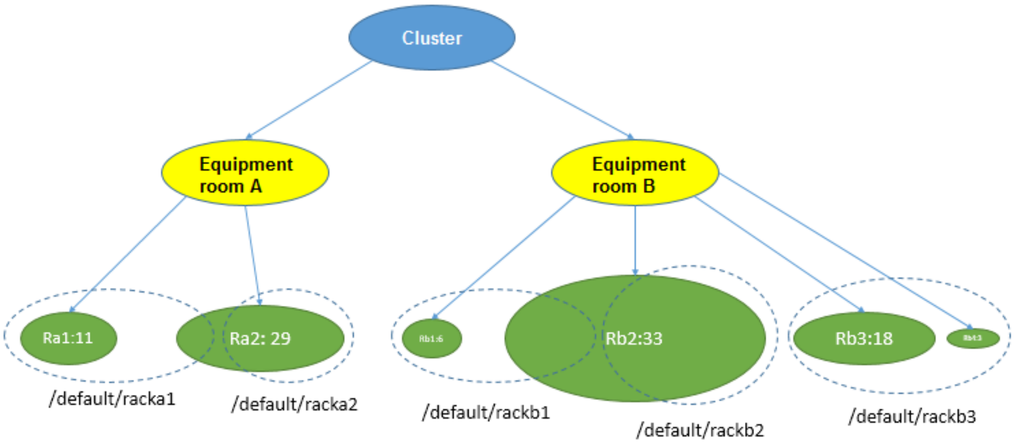
Best Practices

For example, in a cluster, 100 hosts are located in two equipment rooms A and B. A has 40 hosts and B has 60 hosts. In room A, there are 11 hosts on physical rack Ra1 and 29 hosts on physical rack Ra2. In room B, there are six hosts on physical rack Rb1, 33 hosts on physical rack Rb2, 18 hosts on physical rack Rb3, and three hosts on physical rack Rb4.

According to the rack allocation policy, each logical rack contains nearly the same number (for example, 20) of hosts. The allocation details are as follows:

- Logical rack /default/racka1: 11 hosts on physical rack Ra1 and nine hosts on physical rack Ra2
- Logical rack /default/racka2: the remaining 20 hosts (except the nine hosts of logical rack /default/racka1) on physical rack Ra2
- Logical rack /default/rackb1: six hosts on physical rack Rb1 and 13 hosts on physical rack Rb2
- Logical rack /default/rackb2: the remaining 20 hosts on physical rack Rb2
- Logical rack /default/rackb3: 18 hosts on physical rack Rb3 and three hosts on physical rack Rb4

Rack allocation example:



Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Click **Hosts**.

Step 3 Select the check box of the target host.

Step 4 Select **Set Rack** from the **More** drop-down list.

- Set rack names in hierarchy based on the actual network topology. Separate racks from different layers using slashes (/).
- Rack naming rules are as follows: */level1/level2/...* The number of levels must be at least 1, and the name cannot be empty. A rack can contain letters, digits, and underscores (_) and cannot exceed 200 characters.

  For example, /default/rack0.
- If the hosts in the rack to be modified contain DataNode instances, ensure that the rack name levels of the hosts where all DataNode instances reside are the same. Otherwise, the configuration fails to be delivered.

Step 5 Click **OK**.

----End

9.4.2.4 Isolating a Host

Scenario

If a host is abnormal or faulty and cannot provide services or affects the cluster performance, you can remove the host from the available node in the cluster temporarily so that the client can access other available nodes.

NOTE

Only non-management nodes can be isolated.

Impact on the System

- After a host is isolated, all role instances on the host will be stopped, and you cannot start, stop, or configure the host and all instances on the host.
- For some services, after a host is isolated, some instances on other nodes do not work, and the service configuration status may expire.
- After a host is isolated, statistics about the monitoring status and indicator data of the host hardware and instances on the host cannot be collected or displayed.
- Retain the default SSH port (22) of the target node. Otherwise, the task described in this section will fail.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Click **Hosts**.

Step 3 Select the check box of the host to be isolated.

Step 4 Select **Isolate** from the **More** drop-down list.

In the displayed dialog box, enter the password of the current login user and click **OK**.

Step 5 In the displayed confirmation dialog box, select "I confirm to isolate the selected hosts and accept possible consequences of service faults." Click **OK**.

Wait until the message "Operation succeeded" is displayed, and click **Finish**.

The host is successfully isolated and **Running Status** is **Isolated**.

Step 6 Log in to the isolated host as user **root** and run the **pkill -9 -u omm** command to stop the processes of user **omm** on the node. Then run the **ps -ef | grep 'container' | grep '${BIGDATA_HOME}' | awk '{print $2}' | xargs -I '{}' kill -9 '{}'** command to find and stop the container process.

Step 7 Cancel the isolation status of the host before using the host if you have rectified the host exception or fault.

On the **Hosts** page, select the isolated host and choose **More** > **Cancel Isolation**.

NOTE

After the isolation is canceled, all role instances on the host are not started by default. To start role instances on the host, select the target host on the Hosts page and choose **More** > **Start All Instances**.

----End

9.4.2.5 Exporting Host Information

Scenario

Administrators can export information about all hosts on FusionInsight Manager.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Click **Hosts**.

Step 3 Specify the status of required hosts in the drop-down list box on the upper right corner, or click **Advanced Search** to specify hosts.

Step 4 Click **Export All**, select **TXT** or **CSV** for **Save As**, and click **OK**.

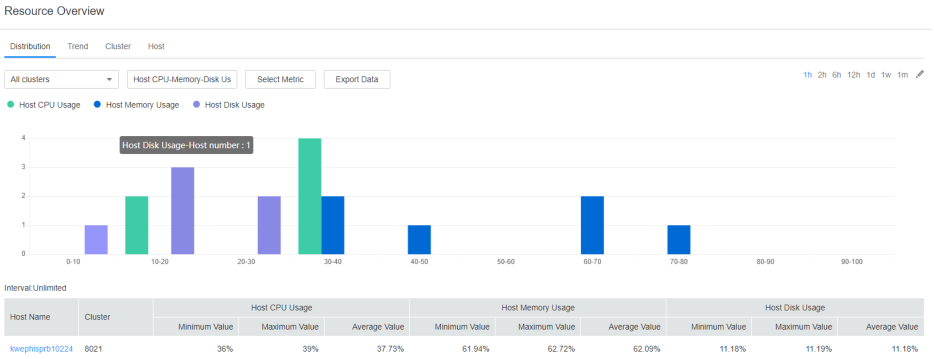
----End

9.4.3 Resource Overview

9.4.3.1 Distribution

Log in to FusionInsight Manager and choose **Hosts** > **Resource Overview**. On the **Resource Overview** page that is displayed, click the **Distribution** tab to view resource distribution of each cluster. By default, the monitoring data of the past one hour (**1h**) is displayed. You can click ✎ to customize a time range. Time range options are **1h**, **2h**, **6h**, **12h**, **1d**, **1w**, and **1m**.

Figure 9-1 Distribution tab



Interval:Unlimited

Host Name	Cluster	Host CPU Usage			Host Memory Usage			Host Disk Usage		
		Minimum Value	Maximum Value	Average Value	Minimum Value	Maximum Value	Average Value	Minimum Value	Maximum Value	Average Value
kwephisprb10224	8021	36%	39%	37.73%	61.94%	62.72%	62.09%	11.18%	11.19%	11.18%

- You can click **Select Metric** to customize the metric to monitor. **Table 9-16** describes all the metrics that you can select. After you select a metric, the host distribution in each range of the metric is displayed.

- When you hover your cursor over a color column, the number of hosts in the current metric range is displayed. See **Figure 9-1**. You can click a color column to view the list of hosts in the metric range.
  - You can click a host name in the **Host Name** column to access the host details page.
  - You can click **View Trends** in the **Operation** column of a host to view the maximum, minimum, and average values of the current metric in the cluster as well as the value of the current host. In the current cluster, if you have selected **Host CPU-Memory-Disk Usage**, **View Trends** is unavailable.
- You can click **Export Data** to export the maximum, minimum, and average values of the current metric of all nodes in the cluster within the time range you have specified.

Table 9-16 Metrics

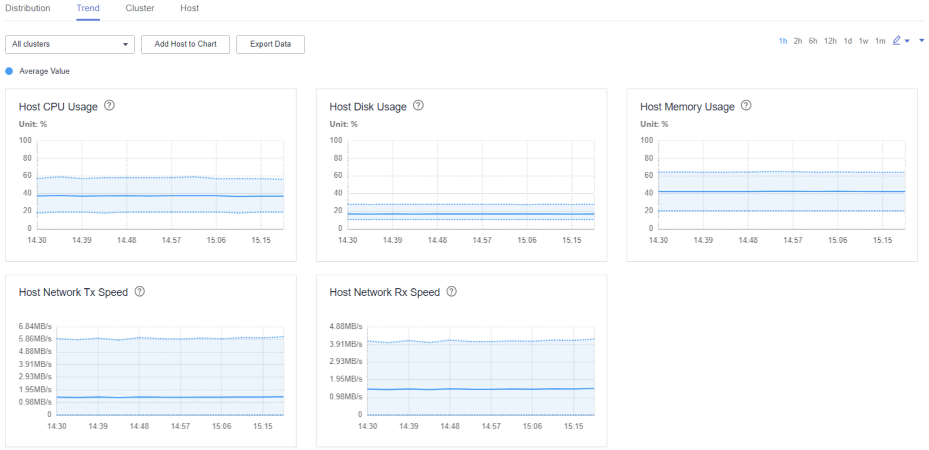
Category	Metric
Process	● Number of Running Processes ● Total Number of Processes ● Total Number of omm Processes ● Uninterruptible Sleep Process
Network Status	● Host Network Packet Collisions ● Number of LAST_ACK States ● Number of CLOSING States ● Number of LISTENING States ● Number of CLOSED States ● Number of ESTABLISHED States ● Number of SYN_RECV States ● Number of TIME_WAITING States ● Number of FIN_WAIT2 States ● Number of FIN_WAIT1 States ● Number of CLOSE_WAIT States ● DNS Name Resolution Duration ● TCP Ephemeral Port Usage ● Host Network Packet Frame Errors
Network Reading	● Host Network Read Packets ● Host Network Read Dropped Packets ● Host Network Read Error Packets ● Host Network Rx Speed

Category	Metric
Disk	● Host Disk Write Speed ● Host Used Disk ● Host Free Disk ● Host Disk Read Speed ● Host Disk Usage
Memory	● Free Memory ● Cache Memory Size ● Total Kernel Cache Memory Size ● Shared Memory Size ● Host Memory Usage ● Used Memory
Network Writing	● Host Network Write Packets ● Host Network Write Error Packets ● Host Network Tx Speed ● Host Network Write Dropped Packets
CPU	● CPU Usage of Processes Whose Priorities Have Been Changed ● CPU Usage of User Space Processes ● CPU Usage of Kernel Space Processes ● Host CPU Usage ● CPU Total Time ● CPU Idle Time
Host Status	● Host File Handle Usage ● Average OS Load in 1 Minute ● Average OS Load in 5 Minutes ● Average OS Load in 15 Minutes ● Host PID Usage

9.4.3.2 Trend

Log in to FusionInsight and choose **Hosts** > **Resource Overview**. On the **Resource Overview** page that is displayed, click the **Trend** tab to view resource trends of all clusters or a single cluster. By default, the monitoring data of the past one hour (**1h**) is displayed. You can click ✎ to customize a time range. Time range options are **1h**, **2h**, **6h**, **12h**, **1d**, **1w**, and **1m**. By default, the trend chart of each metric displays the maximum, minimum, and average values of the entire cluster.

Figure 9-2 Trend tab



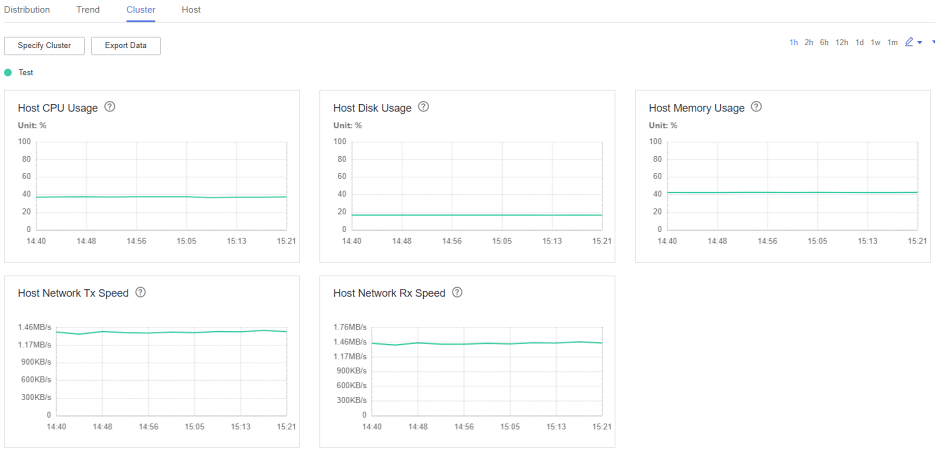
- You can click **Add Host to Chart** to add trend lines of up to 12 hosts to the trend charts.
- You can choose ▾ > **Customize** to customize the metrics to display on the tab page. For details about the metrics, see **Table 9-16** in **Distribution**.
- You can click **Export Data** to export the maximum, minimum, and average values of all nodes in the cluster for all selected metrics within the time range you have specified.

9.4.3.3 Cluster

Log in to FusionInsight Manager and choose **Hosts** > **Resource Overview**. On the **Resource Overview** page that is displayed, click the **Cluster** tab to view resource monitoring of all clusters.

By default, the monitoring data of the past one hour (**1h**) is displayed. You can click ✎ to customize a time range. Time range options are **1h**, **2h**, **6h**, **12h**, **1d**, **1w**, and **1m**.

Figure 9-3 Cluster tab



- You can click **Specify Cluster** to customize a cluster to display.

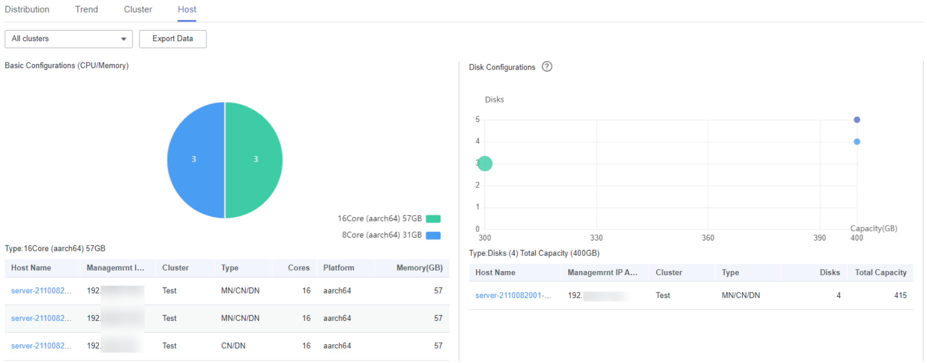
- You can choose ▼ > **Customize** to customize the metrics to display on the tab page. For details about the metrics, see **Table 9-16** in **Distribution**.
- You can click **Export Data** to export the metric values of each cluster within the time range you have specified.

9.4.3.4 Host

Log in to FusionInsight Manager and choose **Hosts** > **Resource Overview**. On the **Resource Overview** page that is displayed, click the **Host** tab to view host resource overview, including basic configurations (CPU/memory) and disk configurations.

You can click **Export Data** to export the configuration list of all hosts in the cluster, including the host name, management IP address, host type, number of cores, CPU architecture, memory capacity, and disk size.

Figure 9-4 Host tab



Type:16Core (aarch64) 57GB

Host Name	Managemrnt I...	Cluster	Type	Cores	Platform	Memory(GB)
server-2110082...	192.[redacted]	Test	MN/CN/DN	16	aarch64	57
server-2110082...	192.[redacted]	Test	MN/CN/DN	16	aarch64	57
server-2110082...	192.[redacted]	Test	CN/DN	16	aarch64	57

Type:Disks (4) Total Capacity (400GB)

Host Name	Managemrnt IP A...	Cluster	Type	Disks	Total Capacity
server-2110082001-...	192.[redacted]	Test	MN/CN/DN	4	415

Basic Configurations (CPU/Memory)

You can hover your cursor over the pie chart to view the number of hosts of each hardware configuration in the cluster. The information is displayed in the format of *Number of cores (CPU architecture) Memory size*.

You can click a slice on the pie chart to view the list of hosts.

Disk Configurations

The horizontal axis indicates the total disk capacity (including the OS disk) of a node, and the vertical axis indicates the number of logical disks (including the OS disk).

You can hover your cursor over a dot to view information about disks of the current configuration, including the quantity of disks, total capacity, and number of hosts.

You can click a dot on the chart to view the list of hosts.

9.5 O&M

9.5.1 Alarms

9.5.1.1 Overview of Alarms and Events

Alarms

Log in to FusionInsight Manager and choose **O&M** > **Alarm** > **Alarms**. You can view information about alarms reported by all clusters, including the alarm name, ID, severity, and generation time. By default, the latest 10 alarms are displayed on each page.

You can click ∨ on the left of an alarm to view detailed alarm parameters. **Table 9-17** describes the parameters.

Table 9-17 Alarm parameters

Parameter	Description
Alarm ID	Alarm ID
Alam Name	Alarm name
Severity	Alarm severity. Value options are **Critical**, **Major**, **Minor**, and **Suggestion**.
Generated	Time when an alarm is generated
Cleared	Time when an alarm is cleared. If the alarm is not cleared, **--** is displayed.
Source	Cluster name
Object	Service, process, or module that triggers the alarm
Automatically Cleared	Whether the alarm can be automatically cleared after the fault is rectified
Alarm Status	Current status of the alarm. Value options are **Auto**, **Manual**, and **Uncleared**.
Alarm Cause	Indicates the possible cause of an alarm.
Serial Number	Indicates the number of alarms generated by the system.
Additional Information	Indicates the error information.

Parameter	Description
Location	Detailed information for locating the alarm, which includes the following: ● **Source**: cluster for which the alarm is generated ● **ServiceName**: service for which the alarm is generated ● **RoleName**: role for which the alarm is generated ● **HostName**: host for which the alarm is generated

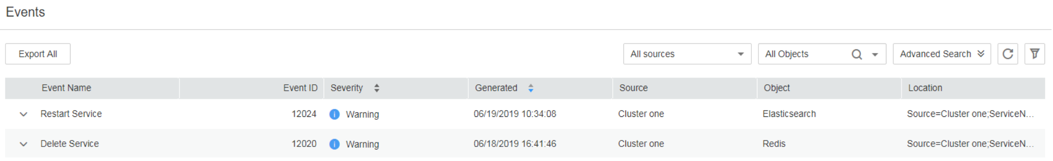
Manage alarms.

- Click **Export All** to export all alarm details.
- If multiple alarms have been handled, you can select one or more alarms to be cleared and click **Clear Alarm** to clear the alarms in batches. A maximum of 300 alarms can be cleared in each batch.
- You can click ↻ to manually refresh the current page and click ⏷ to filter columns to display.
- You can filter alarms by object or cluster.
- You can click **Advanced Search** to search for alarms by alarm ID, name, type, severity, start time, or end time. Click **Search** to filter alarms that meet the search criteria. Click **Advanced Search** again to view the number of search criteria that you have configured.
- You can click **Clear**, **Mask**, or **View Help** to perform corresponding operations on an alarm.
- If there are a large number of alarms, you can click **View by Category** to sort uncleared alarms by alarm ID. After alarms are classified, click the number of uncleared alarms to view alarm details.

Events

Log in to FusionInsight Manager and choose **O&M** > **Alarm** > **Events**. On the **Events** page that is displayed, you can view information about all events in the cluster, including the event name, ID, severity, generation time, object, and location. By default, the latest 10 events are displayed on each page.

Figure 9-5 Events page



Events

Event Name	Event ID	Severity	Generated	Source	Object	Location
Restart Service	12024	Warning	06/19/2019 10:34:08	Cluster one	Elasticsearch	Source=Cluster one;ServiceN...
Delete Service	12020	Warning	06/18/2019 16:41:46	Cluster one	Redis	Source=Cluster one;ServiceN...

You can click ⌄ on the left of an event to view detailed event parameters. Table 9-18 describes the parameters.

Table 9-18 Event parameters

Parameter	Description
Event ID	Event ID
Event Name	Event name
Severity	Event severity. Value options are **Critical**, **Major**, **Minor**, and **Suggestion**.
Generated	Time when an event is generated
Object	Object for which the event may be generated
Serial Number	Number of the event generated by the system
Location	Detailed information for locating the event, which includes the following: ● **Source**: cluster for which the event is generated ● **ServiceName**: service for which the event is generated ● **RoleName**: role for which the event is generated ● **HostName**: host for which the event is generated
Additional Information	Indicates the error information.
Event Cause	Indicates the possible cause of an event.
Source	Cluster name

Manage events.

- Click **Export All** to export all event details.
- You can click ↻ to manually refresh the current page and click ▽ to filter columns to display.
- You can filter events by object or cluster.
- You can click **Advanced Search** to search for events by event ID, name, severity, start time, or end time.

9.5.1.2 Configuring the Threshold

Scenario

You can configure monitoring indicator thresholds to monitor the health status of indicators on FusionInsight Manager. If abnormal data occurs and the preset conditions are met, the system triggers an alarm and displays the alarm information on the alarm page.

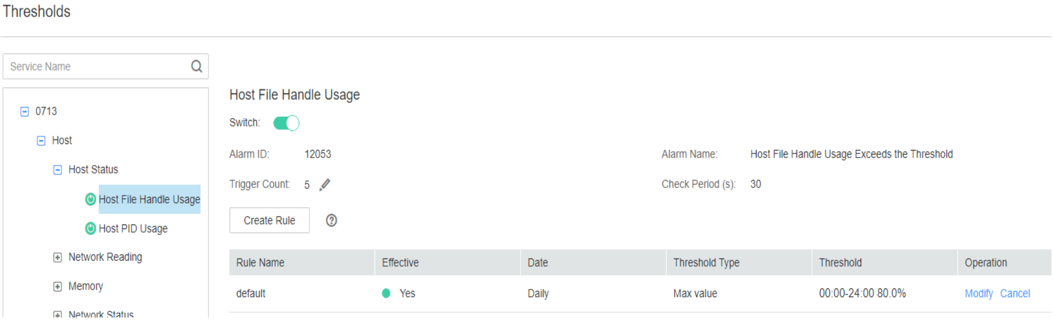
Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **O&M** > **Alarm** > **Thresholds**.

Step 3 Select a monitoring metric for a host or service in the cluster.

Figure 9-6 Configuring the threshold for a metric



Host File Handle Usage

Switch: ●

Alarm ID: 12053

Alarm Name: Host File Handle Usage Exceeds the Threshold

Trigger Count: 5

Check Period (s): 30

Rule Name	Effective	Date	Threshold Type	Threshold	Operation
default	● Yes	Daily	Max value	00:00-24:00 80.0%	Modify Cancel

For example, after selecting **Host Memory Usage**, the information about this indicator threshold is displayed.

- If the alarm sending switch is displayed as ●, an alarm is triggered if the threshold is reached.
- **Alarm ID** and **Alarm Name**: alarm information triggered against the threshold
- **Trigger Count**: FusionInsight Manager checks whether the value of a monitoring metric reaches the threshold. If the number of consecutive checks reaches the value of **Trigger Count**, an alarm is generated. **Trigger Count** is configurable.
- **Check Period (s)**: interval for the system to check the monitoring metric.
- The rules in the rule list are used to trigger alarms.

Step 4 Click **Create Rule** to add rules used for monitoring indicators.

Table 9-19 Monitoring indicator rule parameters

Parameter	Description	Example Value
Rule Name	Name of a rule.	CPU_MAX
Severity	Alarm Severity ● Critical ● Major ● Minor ● Warning	● Critical ● Major ● Minor ● Warning

Parameter	Description	Example Value
Threshold Type	You can use the maximum or minimum value of an indicator as the alarm triggering threshold. If **Threshold Type** is set to **Max value**, the system generates an alarm when the value of the specified indicator is greater than the threshold. If **Threshold Type** is set to **Min value**, the system generates an alarm when the value of the specified indicator is less than the threshold.	● Max value ● Min value
Date	This parameter is used to set the date when the rule takes effect.	● Daily ● Weekly ● Others
Add Date	This parameter is available only when **Date** is set to **Others**. You can set the date when the rule takes effect. Multiple options are available.	09-30
Thresholds	This parameter is used to set the time range when the rule takes effect.	Start and End Time: 00:00–08:30
	Threshold of the rule monitoring metric	Threshold: 10

NOTE

You can click ⊕ or 🗑 to add or delete time thresholds.

Step 5 Click **OK** to save the rules.

Step 6 Locate the row that contains an added rule, and click **Apply** in the **Operation** column. The value of **Effective** for this rule changes to **Yes**.

A new rule can be applied only after you click **Cancel** for an existing rule.

----End

Monitoring Metric Reference

FusionInsight Manager alarm monitoring metrics are classified as node information metrics and cluster service metrics. **Table 9-20** describes the metrics for which you can configure thresholds on nodes.

Table 9-20 Node monitoring metrics

Metric Group	Metric	Description	Default Threshold
CPU	Host CPU Usage	This indicator reflects the computing and control capabilities of the current cluster in a measurement period. By observing the indicator value, you can better understand the overall resource usage of the cluster.	90.0%
Disk	Disk Usage	Indicates the disk usage of a host.	90.0%
	Disk Inode Usage	Indicates the disk inode usage in a measurement period.	80.0%
Memory	Host Memory Usage	Indicates the average memory usage at the current time.	90.0%
Host Status	Host File Handle Usage	Indicates the usage of file handles of the host in a measurement period.	80.0%
	Host PID Usage	Indicates the PID usage of a host.	90%

Metric Group	Metric	Description	Default Threshold
Network Status	TCP Ephemeral Port Usage	Indicates the usage of temporary TCP ports of the host in a measurement period.	80.0%
Network Reading	Read Packet Error Rate	Indicates the read packet error rate of the network interface on the host in a measurement period.	0.5%
	Read Packet Dropped Rate	Indicates the read packet dropped rate of the network interface on the host in a measurement period.	0.5%
	Read Throughput Rate	Indicates the average read throughput (at MAC layer) of the network interface in a measurement period.	80%
Network Writing	Write Packet Error Rate	Indicates the write packet error rate of the network interface on the host in a measurement period.	0.5%
	Write Packet Dropped Rate	Indicates the write packet dropped rate of the network interface on the host in a measurement period.	0.5%

Metric Group	Metric	Description	Default Threshold
	Write Throughput Rate	Indicates the average write throughput (at MAC layer) of the network interface in a measurement period.	80%
Process	Uninterruptible Sleep Process	Number of D state processes on the host in a measurement period	0
	omm Process Usage	omm process usage in a measurement period	90

Table 9-21 Cluster service indicators

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
DBService	Database	Usage of the Number of Database Connections	Indicates the usage of the number of database connections.	90%
		Disk Space Usage of the Data Directory	Disk space usage of the data directory	80%
Flume	Agent	Heap Memory Usage Calculate	Indicates the Flume heap memory usage.	95.0%
		Flume Direct Memory Usage Statistics	Indicates the Flume direct memory usage.	80.0%
		Flume Non-heap Memory Usage	Indicates the Flume non-heap memory usage.	80.0%

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
		Total GC duration of Flume process	Indicates the Flume total GC time.	12000 ms
HBase	GC	GC time for old generation	Total GC time of RegionServer	5000 ms
		GC time for old generation	Indicates he total GC time of HMaster.	5000 ms
	CPU & memory	RegionServer Direct Memory Usage Statistics	Indicates theRegionServerReg direct memory usage.	90%
		RegionServer Heap Memory Usage Statistics	Indicates the RegionServer heap memory usage.	90%
		HMaster Direct Memory Usage	Indicates the HMaster direct memory usage.	90%
		HMaster Heap Memory Usage Statistics	Indicates the HMaster heap memory usage.	90%
	Service	Number of Online Regions of a RegionServer	Number of regions of a RegionServer	2000
		Region in transaction count over threshold	Number of regions that are in the RIT state and reach the threshold duration	1

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
	Replication	Replication sync failed times (RegionServer)	Indicates the number of times that DR data fails to be synchronized.	1
		Number of Log Files to Be Synchronized in the Active Cluster	Number of log files to be synchronized in the active cluster	128
		Number of HFiles to Be Synchronized in the Active Cluster	Number of HFiles to be synchronized in the active cluster	128
	Queue	Compaction Queue Size	Size of the Compaction queue	100
HDFS	File and Block	Lost Blocks	Indicates the number of block copies that the HDFS lacks of.	0
		Blocks Under Replicated	Total number of blocks that need to be replicated by the NameNode	1000
	RPC	Average Time of Active NameNode RPC Processing	Indicates the average RPC processing time.	100 ms
		Average Time of Active NameNode RPC Queuing	Indicates the average RPC queuing time.	200 ms
	Disk	HDFS Disk Usage	Indicates the HDFS disk usage.	80%

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
		DataNode Disk Usage	Indicates the disk usage of DataNodes in the HDFS.	80%
		Percentage of Reserved Space for Replicas of Unused Space	Indicates the percentage of the reserved disk space of all the copies to the total unused disk space of DataNodes.	90%
	Resource	Faulty DataNodes	Indicates the number of faulty DataNodes.	3
		NameNode Non Heap Memory Usage Statistics	Indicates the percentage of NameNode non-heap memory usage.	90%
		NameNode Direct Memory Usage Statistics	Indicates the percentage of direct memory used by NameNodes.	90%
		NameNode Heap Memory Usage Statistics	Indicates the percentage of NameNode non-heap memory usage.	95%
		DataNode Direct Memory Usage Statistics	Indicates the percentage of direct memory used by DataNodes.	90%
		DataNode Heap Memory Usage Statistics	DataNode heap memory usage	95%

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
		DataNode Heap Memory Usage Statistics	Indicates the percentage of DataNode non-heap memory usage.	90%
	Garbage Collection	GC Time (NameNode)/ GC Time (DataNode)	Indicates the Garbage collection (GC) duration of NameNodes per minute.	12000 ms
		GC Time	Indicates the GC duration of DataNodes per minute.	12000 ms
Hive	HQL	Percentage of HQL Statements That Are Executed Successfully by Hive	Indicates the percentage of HQL statements that are executed successfully by Hive.	90.0%
	Background	Background Thread Usage	Background thread usage	90%
	GC	Total GC time of MetaStore	Indicates the total GC time of MetaStore.	12000 ms
		Total GC Time in Milliseconds	Indicates the total GC time of HiveServer.	12000 ms
	Capacity	Percentage of HDFS Space Used by Hive to the Available Space	Indicates the percentage of HDFS space used by Hive to the available space.	85.0%

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
	CPU & memory	MetaStore Direct Memory Usage Statistics	MetaStore direct memory usage	95%
		MetaStore Non-Heap Memory Usage Statistics	MetaStore non-heap memory usage	95%
		MetaStore Heap Memory Usage Statistics	MetaStore heap memory usage	95%
		HiveServer Direct Memory Usage Statistics	HiveServer direct memory usage	95%
		HiveServer Non-Heap Memory Usage Statistics	HiveServer non-heap memory usage	95%
		HiveServer Heap Memory Usage Statistics	HiveServer heap memory usage	95%
	Session	Percentage of Sessions Connected to the HiveServer to Maximum Number of Sessions Allowed by the HiveServer	Indicates the percentage of the number of sessions connected to the HiveServer to the maximum number of sessions allowed by the HiveServer.	90.0%

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
Kafka	Partition	Percentage of Partitions That Are Not Completely Synchronized	Indicates the percentage of partitions that are not completely synchronized to total partitions.	50%
	Others	Unavailable Partition Percentage	Percentage of unavailable partitions of each Kafka topic	40%
		User Connection Usage on Broker	Usage of user connections on Broker	80%
	Disk	Broker Disk Usage	Indicates the disk usage of the disk where the Broker data directory is located.	80.0%
		Disk I/O Rate of a Broker	I/O usage of the disk where the Broker data directory is located	80%
	Process	Broker GC Duration per Minute	Indicates the GC duration of the Broker process per minute.	12000 ms
		Heap Memory Usage of Kafka	Indicates the Kafka heap memory usage.	95%
		Kafka Direct Memory Usage	Indicates the Kafka direct memory usage.	95%

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
Loader	Memory	Heap Memory Usage Calculate	Indicates the Loader heap memory usage.	95%
		Direct Memory Usage of Loader	Indicates the Loader direct memory usage.	80.0%
		Non-heap Memory Usage of Loader	Indicates the Loader non-heap memory usage.	80%
	GC	Total GC time of Loader	Indicates the total GC time of Loader.	12000 ms
MapReduce	Garbage Collection	GC Time	Indicates the GC time.	12000 ms
	Resource	JobHistoryServer Direct Memory Usage Statistics	Indicates the JobHistoryServer direct memory usage.	90%
		JobHistoryServer Non Heap Memory Usage Statistics	Indicates the JobHistoryServer non-heap memory usage.	90%
		JobHistoryServer Heap Memory Usage Statistics	Indicates the JobHistoryServer non-heap memory usage.	95%
Oozie	Memory	Heap Memory Usage Calculate	Indicates the Oozie heap memory usage.	95.0%
		Oozie Direct Memory Usage	Indicates the Oozie direct memory usage.	80.0%

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
		Oozie Non-heap Memory Usage	Indicates the Oozie non-heap memory usage.	80%
	GC	Total GC duration of Oozie	Indicates the Oozie total GC time.	12000 ms
Spark2x	Memory	JDBCServer2x Heap Memory Usage Statistics	JDBCServer2x heap memory usage	95%
		JDBCServer2x Direct Memory Usage Statistics	JDBCServer2x direct memory usage	95%
		JDBCServer2x Non-Heap Memory Usage Statistics	JDBCServer2x non-heap memory usage	95%
		JobHistory2x Direct Memory Usage Statistics	JobHistory2x direct memory usage	95%
		JobHistory2x Non-Heap Memory Usage Statistics	JobHistory2x non-heap memory usage	95%
		JobHistory2x Heap Memory Usage Statistics	JobHistory2x heap memory usage	95%
		IndexServer2x Direct Memory Usage Statistics	IndexServer2x direct memory usage	95%

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
		IndexServer2x Heap Memory Usage Statistics	IndexServer2x heap memory usage	95%
		IndexServer2x Non-Heap Memory Usage Statistics	IndexServer2x non-heap memory usage	95%
	GC Count	Full GC Number of JDBCServer2x	Total GC number of JDBCServer2x	12
		Full GC Number of JobHistory2x	Total GC number of JobHistory2x	12
		Full GC Number of IndexServer2x	Total GC number of IndexServer2x	12
	GC Time	Total GC Time in Milliseconds	Total GC time of JDBCServer2x	12000 ms
		Total GC Time in Milliseconds	Total GC time of JobHistory2x	12000 ms
		Total GC Time in Milliseconds	Total GC time of IndexServer2x	12000 ms
Storm	Cluster	Number of Available Supervisors	Indicates the number of available Supervisor processes in the cluster in a measurement period.	1
		Slot Usage	Indicates the slot usage in the cluster in a measurement period.	80.0%

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
	Nimbus	Heap Memory Usage Calculate	Indicates the Nimbus heap memory usage.	80%
Yarn	Resources	NodeManager Direct Memory Usage Statistics	Indicates the percentage of direct memory used by NodeManagers.	90%
		NodeManager Heap Memory Usage Statistics	Indicates the percentage of NodeManager heap memory usage.	95%
		NodeManager Non Heap Memory Usage Statistics	Indicates the percentage of NodeManager non-heap memory usage.	90%
		ResourceManager Direct Memory Usage Statistics	Indicates the Kafka direct memory usage.	90%
		ResourceManager Heap Memory Usage Statistics	Indicates the ResourceManager heap memory usage.	95%
		ResourceManager Non Heap Memory Usage Statistics	Indicates the ResourceManager non-heap memory usage.	90%
	Garbage collection	GC Time	Indicates the GC duration of NodeManager per minute.	12000 ms

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
		GC Time	Indicates the GC duration of ResourceManager per minute.	12000 ms
	Others	Failed Applications of root queue	Number of failed tasks in the root queue	50
		Terminated Applications of root queue	Number of killed tasks in the root queue	50
	CPU & memory	Pending Memory	Pending memory capacity	83886080MB
	Application	Pending Applications	Pending tasks	60
ZooKeeper	Connection	ZooKeeper Connections Usage	Indicates the percentage of the used connections to the total connections of ZooKeeper.	80%
	CPU & memory	Directmemory Usage Calculate	Indicates the ZooKeeper heap memory usage.	95%
		Heap Memory Usage Calculate	Indicates the ZooKeeper direct memory usage.	80%
	GC	ZooKeeper GC Duration per Minute	Indicates the GC time of ZooKeeper every minute.	12000 ms
meta	OBS data write operation	Success Rate for Calling the OBS Write API	Success rate for calling the OBS data read API	99.0%

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
	OBS Meta data Operations	Average Time for Calling the OBS Metadata API	Average time for calling the OBS metadata API	500ms
		Success Rate for Calling the OBS Metadata API	Success rate for calling the OBS metadata API	99.0%
	OBS data read operation	Success Rate for Calling the OBS Data Read API	Success rate for calling the OBS data read API	99.0%
Ranger	GC	UserSync GC Duration	UserSync garbage collection (GC) duration	12000 ms
		RangerAdmin GC Duration	RangerAdmin GC duration	12000 ms
		TagSync GC Duration	TagSync GC duration	12000 ms
	CPU & memory	UserSync Non-Heap Memory Usage	UserSync non-heap memory usage	80.0%
		UserSync Direct Memory Usage	UserSync direct memory usage	80.0%
		UserSync Heap Memory Usage	UserSync heap memory usage	95.0%
		RangerAdmin Non-Heap Memory Usage	RangerAdmin non-heap memory usage	80.0%
		RangerAdmin Heap Memory Usage	RangerAdmin heap memory usage	95.0%

Service	Monitoring Indicator Group Name	Indicator Name	Description	Default Threshold
		RangerAdmin Direct Memory Usage	RangerAdmin direct memory usage	80.0%
		TagSync Direct Memory Usage	TagSync direct memory usage	80.0%
		TagSync Non-Heap Memory Usage	TagSync non-heap memory usage	80.0%
		TagSync Heap Memory Usage	TagSync heap memory usage	95.0%
ClickHouse	Cluster Quota	Clickhouse service quantity quota usage in ZooKeeper	Quota of the ZooKeeper nodes used by a ClickHouse service	90%
		Capacity quota usage of the Clickhouse service in ZooKeeper	Capacity quota of ZooKeeper directory used by the ClickHouse service	90%

9.5.1.3 Configuring the Alarm Masking Status

Scenario

If you do not want FusionInsight Manager to report specified alarms in the following scenarios, you can manually mask the alarms.

- Some unimportant alarms and minor alarms need to be masked.
- When a third-party product is integrated with FusionInsight, some alarms of the product are duplicated with the alarms of FusionInsight and need to be masked.
- When the deployment environment is special, certain alarms may be falsely reported and need to be masked.

After an alarm is masked, new alarms with the same ID as the alarm are neither displayed on the **Alarm** page nor counted. The reported alarms are still displayed.

Procedure

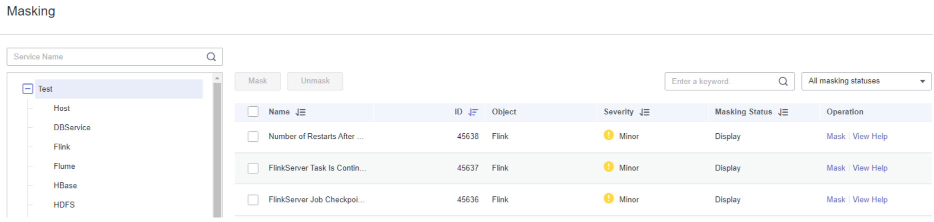
Step 1 Log in to FusionInsight Manager.

Step 2 Choose **O&M** > **Alarm** > **Masking Setting**.

Step 3 In the **Masking Setting** area, select the specified service or module.

Step 4 Select an alarm from the alarm list.

Figure 9-7 Masking an alarm



□	Name	ID	Object	Severity	Masking Status	Operation
□	Number of Restarts After ...	45638	Flink	Minor	Display	Mask | View Help
□	FlinkServer Task Is Contin...	45637	Flink	Minor	Display	Mask | View Help
□	FlinkServer Job Checkpoi...	45636	Flink	Minor	Display	Mask | View Help

The information about the alarm is displayed, including the alarm name, ID, severity, masking status, and operations can be performed on the alarm.

- The masking status includes **Display** and **Masking**.
- Operations include **Masking** and **Help**.

NOTE

You can filter specified alarms based on the masking status and alarm severity.

Step 5 Set the masking status for an alarm:

- Click **Masking**. In the displayed dialog box, click **OK** to change the alarm masking status to **Masking**.
- Click **Cancel Masking**. In the dialog box that is displayed, click **OK** to change the masking status of the alarm to **Display**.

----End

9.5.2 Log

9.5.2.1 Log Online Search

Scenario

FusionInsight Manager allows you to search for logs online and view the log content of components to locate faults.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **O&M** > **Log** > **Online Search**.

Step 3 Configure the parameters listed in Table 9-22 to search for the logs you need. You can select a default log search duration (including **0.5h**, **1h**, **2h**, **6h**, **12h**, **1d**, **1w**, and **1m**), or click ✎ to customize **Start Data** and **End Data**.

Table 9-22 Log search parameters

Parameter	Description
Search Content	Keywords or regular expression to be searched for
Service	Service or module for which you want to query logs
File	Log files to be searched for when only one role is selected
Lowest Log Level	Lowest level of logs to be queried. After you select a level, the logs of this level and higher levels are displayed. The levels in ascending order are as follows: TRACE < DEBUG < INFO < WARN < ERROR < FATAL
Host Scope	● You can click ... to select hosts. ● Enter the host name of the node for which you want to query logs or the IP address of the management plane. ● Use commas (,) to separate IP addresses, for example, **192.168.10.10**,**192.168.10.11**. ● Use hyphens (-) to indicate an IP address segment if the IP addresses are consecutive, for example, **192.168.10.[10-20]**. ● Use hyphens (-) to indicate an IP address segment if the IP addresses are consecutive, and use commas (,) to separate IP address segments, for example, **192.168.10.[10-20,30-40]**. **NOTE** – If this parameter is not specified, all hosts are selected by default. – A maximum of 10 expressions can be entered at a time. – A maximum of 2,000 hosts can be matched for all entered expressions at a time.
Advanced Configurations	● **Max Quantity**: maximum number of logs that can be displayed at a time. If the number of queried logs exceeds the value of this parameter, the earliest logs will be ignored. If this parameter is not set, the maximum number of logs that can be displayed at a time is not limited. ● **Timeout Duration**: log query timeout duration. This parameter is used to limit the maximum log query time on each node. When the query times out, the query is stopped and the logs that have been searched for are still displayed.

Step 4 Click **Search**. **Table 9-23** describes the fields in search results.

Table 9-23 Parameters in search results

Parameter	Description
Time	Time when a line of log is generated
Source Cluster	Cluster for which the log is generated
Host Name	Host name of the node where the log file recording the line of log is located
Location	Path of the log file recording the line of log Click the location information to go to the online log browsing page. By default, 100 lines of logs before and 100 lines after the line of log are displayed. You can click **Load More** on the top or bottom of the page to view more logs. Click **Download** to download the log file to the local PC.
Line No.	Line number of a line of log in the log file
Level	Level of the line of log
Log	Log content

NOTE

You can click **Stop** to forcibly stop the search. You can view the search results in the list.

Step 5 Click **Filter** to filter the logs to display on the page. **Table 9-24** lists the fields that you can use to filter logs. After you configure these parameters, click **Filter** to search for logs meeting the search criteria. You can click **Reset** to clear the information that you have filled in.

Table 9-24 Parameters for filtering logs

Parameter	Description
Keywords	Keywords of the losg to be searched for
Host Name	Name of the host to be searched for
Location	Path of the log file to be searched for
Started	Start time for logs to be searched for
Completed	End time for logs to be searched for
Source Cluster	Cluster for which logs need to be searched

----End

9.5.2.2 Log Download

Scenario

FusionInsight Manager allows you to batch export logs generated on all instances of each service.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **O&M** > **Log** > **Download**.

Step 3 Select a log download range:

1. **Service**: Click ··· and select a service.
2. **Host**: Enter the IP address of the host where the service is deployed. You can also click ··· to select the required host.
3. Click ✎ in the upper right corner and configure **Start Time** and **End Time**.

Step 4 Click **Download**.

The downloaded log package contains the topology information of the start time and end time, helping you quickly find the log you need.

The topology file is named in the format of **topo_<***Topology structure change time***>.txt**. The file contains the node IP address, host name, and service instances that reside on the node. (OMS nodes are identified by **Manager:Manager**.)

Example:

```
192.168.204.124|suse-124|
DBService:DBServer;KrbClient:KerberosClient;LdapClient:SlapdClient;LdapServer:SlapdServer;Manager:Manag
er;meta:meta
```

----End

9.5.3 Perform a Health Check

9.5.3.1 Viewing a Health Check Task

Scenario

Administrators can view all health check tasks in the health check management center to check whether the cluster is affected after the modification.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **O&M** > **Health Check**.

By default, all saved health check reports are listed. The parameters for a health check report are as follows:

Table 9-25 Parametes for a health check report

Parameter	Description
Check Object	Object to be checked. You can expand the list to view its details.
Status	Check result status. Value options are **No problems found**, **Problems found**, and **Checking**.
Check Type	Entity on which the check is to be performed. Value options are **System**, **Cluster**, **Host**, **Service**, and **OMS**. If you select **Cluster**, all items are checked by default.
Start Mode	Whether the health check is automatically or manually performed
Started	Start time of the check
Completed	End time of the check
Operation	Operations you can perform. Value options are **Export Report** and **View Help**.

NOTE

- In the upper right corner of the check list, you can filter health checks by check type or status.
- If **Check Type** is **Cluster**, **View Help** is displayed in the **Check Object** drop-down list.
- During a health check, the system determines whether check objects are healthy based on their historical monitoring metric data.

----End

9.5.3.2 Managing Health Check Reports

Scenario

FusionInsight Manager allows you to download and delete health check reports.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **O&M** > **Health Check**.

Step 3 Locate the row containing the target health check report and click **Export Report** in the **Operation** to download the report.

----End

9.5.3.3 Modifying Health Check Configuration

Scenario

Administrators can enable automatic health check to reduce manual operation time. By default, the automatic health check checks the entire cluster.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **O&M** > **Health Check** > **Configuration**.

Periodic Health Check indicates whether to enable automatic health check. Selecting **Enable** to enable the automatic health check, and selecting **Disable** to disable the function.

Set the health check period to **Daily**, **Weekly**, or **Monthly** as required.

Step 3 Click **OK** to save the configurations.

----End

9.5.4 Configuring Backup and Backup Restoration

9.5.4.1 Creating a Backup Task

Scenario

You can create backup tasks on FusionInsight Manager. Executing backup tasks backs up related data.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **O&M** > **Backup and Restoration** > **Backup Management**. On the page that is displayed, click **Create**.

Step 3 Set **Backup Object** to **OMS** or the cluster whose data you want to back up.

Step 4 Enter a task name in the **Name** text box.

Step 5 Set **Mode** to **Periodic** or **Manual** as required.

Table 9-26 Backup types

Type	Parameter	Description
Periodic backup	Start Time	Indicates the time when a periodic backup task is started for the first time.
	Period	Task execution interval. Value options are **Hours** and **Days**.

Type	Parameter	Description
	Backup Policy	The following policies can be selected: ● Full backup at the first time and subsequent incremental backup ● Full backup every time ● Full backup once every n times
Manual backup	N/A	You need to manually execute the task to back up data.

Step 6 Set required parameters in the **Configuration** area.

- Metadata and service data can be backed up.
- For details about how to back up data of different components, see **Backup and Recovery Management**.

Step 7 Click **OK** to save the configurations.

Step 8 In the backup task list, you can view the created backup task.

Locate the row that contains the target backup task, choose **More** > **Back Up Now** in the **Operation** column to execute the task immediately.

----End

9.5.4.2 Creating a Backup Restoration Task

Scenario

You can create a backup restoration task on FusionInsight Manager. After the restoration task is executed, the specified backup data is restored to the cluster.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **O&M** > **Backup and Restoration** > **Restoration Management**. On the page that is displayed, click **Create**.

Step 3 Configure **Task Name**.

Step 4 Set **Recovery Object** to **OMS** or the cluster whose data you want to restore.

Step 5 Set the required parameters in the **Recovery Configuration** area.

- Metadata and service data can be restored.
- For details about how to restore data of different components, see **Backup and Recovery Management**.

Step 6 Click **OK** to save the configurations.

Step 7 In the restoration task list, you can view the created restoration tasks.

Locate the row containing the target restoration task, click **Start** in the **Operation** column to execute the restoration task immediately.

----End

9.5.4.3 Managing Backup and Backup Restoration Tasks

Scenario

You can also maintain and manage backup restoration tasks on FusionInsight Manager.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **O&M** > **Backup and Restoration** > **Backup Management** or **Restoration Management**.

Step 3 In the **Operation** column of the specified task in the task list, select the operation to be performed.

Table 9-27 Maintenance and management operations

Operation Entry	Description
Config	Modify parameters for the backup task.
Recover	After some service data is successfully backed up, you can use this function to quickly restore data.
More > **Back Up Now**	Perform this operation to execute the backup task immediately.
More > **Stop**	Perform this operation to stop a running task.
More > **Delete** or **Delete**	This operation is used to delete tasks.
More > **Suspend**	Perform this operation to disable the automatic backup task function.
More > **Resume**	Perform this operation to enable the automatic backup task function.
More > **View History** or **View History**	Perform this operation to switch to the task run log page to view the task running details and backup path.
View	Perform this operation to check the parameter settings of the restoration task.
Start	Perform this operation to run the restoration task.

----End

9.6 Audit

9.6.1 Overview

Scenario

The **Audit** page displays the user operations on Manager. On this page, administrators can view historical user operations on Manager. For details about the audit information, see **Audit Logs**.

Overview

Log in to FusionInsight Manager and choose **Audit**. The **Audit** page displays the operation type, risk level, start time, end time, user, source, host name, service, instance, and operation result.

- You can select audit logs at the **Critical**, **Major**, **Minor**, or **Notice** level from the **All risk levels** drop-down list.
- In **Advanced Search**, you can set filter criteria to query audit logs.
  a. You can query audit logs by user management, cluster, service, and health in the **Operation Type** column.
  b. In the **Service** column, you can select a service to query corresponding audit logs.

     NOTE

     You can select **--** to search for audit logs using all other search criteria except services.
  c. You can query audit logs by operation result. Value options are **All**, **Successful**, **Failed**, and **Unknown**.
- You can click ↻ to manually refresh the current page or click ▽ to filter the columns displayed in the page.
- Click **Export All** to export all audit information at a time. The audit information can be exported in **TXT** or **CSV** format.

9.6.2 Configuring Audit Log Dumping

Scenario

The audit logs of FusionInsight Manager are stored in the database by default. If the audit logs are retained for a long time, the disk space of the data directory may be insufficient. To store audit logs to another archive server, administrators can set the required dump parameters to automatically dump these logs. This facilitates the management of audit logs.

If you do not configure the audit log dumping, the system automatically saves the audit logs to a file when the number of audit logs reaches 100,000 pieces. The save path is **${BIGDATA_DATA_HOME} /dbdata_om/dumpData/iam/operatelog** on the active management node. The file name format is

OperateLog_store_*YY_MM_DD_HH_MM_SS***.csv**. The maximum number of historical audit log files is 50.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Audit** > **Configuration**.

Step 3 Click the switch on the right of **Audit Log Dumping Flag**.

Audit Log Dump is disabled by default. If ● is displayed, **Audit Log Dump** is enabled.

Step 4 Set the dump parameters based on information provided in **Table 9-28**

Table 9-28 Audit log dump parameters

Parameter	Description	Value
SFTP IP Mode	Mode of the destination IP address. The value can be **IPv4** or **IPv6**.	IPv4
SFTP IP	SFTP server for storing dumped audit logs. You are advised to use the SFTP service based on SSH v2 to prevent security risks.	**192.168.10.51** (example value)
SFTP Port	Connection port of the SFTP server for storing dumped audit logs	**22** (example value)
Save Path	Path for storing audit logs on the SFTP server	**/opt/omm/oms/auditLog** (example value)
SFTP Username	User name for logging in to the SFTP server	**root** (example value)
SFTP Password	Password for logging in to the SFTP server	*Password for logging into the SFTP server*
SFTP Public key	Specifies the public key of the SFTP server. This parameter is optional. You are advised to set the public key of the SFTP server. Otherwise, security risks may exist.	-

Parameter	Description	Value
Dumping Mode	Dump mode. Value options are as follows: ● **By Quantity**: If the number of pieces of logs reaches the value of this parameter (**100000** by default), the logs are dumped. ● **By Time**: specifies the date when logs are dumped. The dumping frequency is once a year.	● By Quantity ● By Time
Dumping Date	This parameter is available only when **Dumping Mode** is set to **By time**. After you select a dump date, the system starts dumping on this date. The logs to be dumped include all the audit logs generated before January 1 00:00 of the current year.	11-06

NOTE

If the SFTP public key is empty, the system displays a security risk warning. Evaluate the security risk and then save the configuration.

Step 5 Click **OK** to complete the settings.

NOTE

Key fields in the audit log dump file are as follows:

- **USERTYPE** indicates the user type. Value **0** indicates a human-machine user, and value **1** indicates a machine-machine user.
- **LOGLEVEL** indicates the security level. Value **0** indicates Critical, value **1** indicates Major, value **2** indicates Minor, and value **3** indicates Warning.
- **OPERATERESULT** indicates the operation result. Value **0** indicates that the operation is successful, and value **1** indicates that the operation is failed.

----End

9.7 Tenant Resources

9.7.1 Multi-Tenancy

9.7.1.1 Overview

Definition

Multi-tenancy refers to multiple resource sets (a resource set is a tenant) in the MRS big data cluster and is able to allocate and schedule resources. The resources include computing resources and storage resources.

Context

Modern enterprises' data clusters are becoming more and more centralized and cloud-based. Enterprise-class big data clusters must meet the following requirements:

- Carry data of different types and formats and run jobs and applications of different types (such analysis, query, and stream processing).
- Isolate data of a user from that of another user who has demanding requirements on data security, such as a bank or government institute.

The preceding requirements bring the following challenges to the big data clusters:

- Proper allocation and scheduling of resources to ensure stable operating of applications and jobs.
- Strict access control to ensure data and service security.

Multi-tenancy isolates the resources of a big data cluster into resource sets. Users can lease desired resource sets to run applications and jobs and store data. In a big data cluster, multiple resource sets can be deployed to meet diverse requirements of multiple users.

The MRS big data cluster provides a complete enterprise-class big data multi-tenant solution.

Highlights

- Proper resource configuration and isolation

  The resources of a tenant are isolated from those of another tenant. The resource use of a tenant does not affect other tenants. This mechanism ensures that each tenant can configure resources based on service requirements, improving resource utilization.
- Resource consumption measurement and statistics

  Tenants are system resource applicants and consumers. System resources are planned and allocated based on tenants. Resource consumption by tenants can be measured and collected.
- Assured data security and access security

  In multi-tenant scenarios, the data of each tenant is stored separately to ensure data security. The access to tenants' resources is controlled to ensure access security.

9.7.1.2 Technical Principles

9.7.1.2.1 Multi-Tenant Management

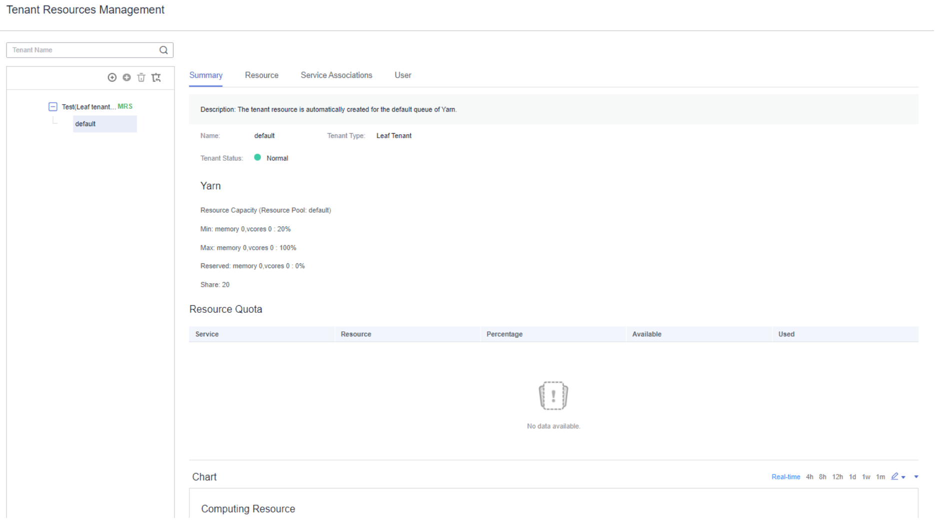
Unified Multi-Tenant Management

Log in to FusionInsight Manager and choose **Tenant Resources** > **Tenant Resources Management**. On the page that is displayed, you can find that FusionInsight Manager is a unified multi-tenant management platform that integrates multiple functions such as tenant lifecycle management, tenant resource configuration, tenant service association, and tenant resource usage statistics, delivering a mature multi-tenant management model and achieving centralized tenant and service management.

Graphical User Interface

FusionInsight Manager provides the graphical multi-tenant management interface and manages and operates multiple levels of tenants using the tree structure. Additionally, FusionInsight Manager integrates the basic information and resource quota of the current tenant in one interface to facilitate O&M and management, as shown in **Figure 9-8**.

Figure 9-8 Tenant management page of FusionInsight Manager



Description: The tenant resource is automatically created for the default queue of Yarn.

Name: default

Tenant Type: Leaf Tenant

Tenant Status: ● Normal

Yarn

Resource Capacity (Resource Pool: default)

Min: memory 0,vcores 0 : 20%

Max: memory 0,vcores 0 : 100%

Reserved: memory 0,vcores 0 : 0%

Share: 20

Resource Quota

Service	Resource	Percentage	Available	Used

No data available.

Chart

Computing Resource

Hierarchical Tenant Management

FusionInsight Manager supports a hierarchical tenant management model in which you can add sub-tenants to an existing tenant to re-configure resources. Sub-tenants of level-1 tenants are level-2 tenants. So on and so forth. FusionInsight Manager provides enterprises with a field-tested multi-tenant management model, enabling centralized tenant and service management.

Simplified Permission Management

FusionInsight Manager hides internal permission management details from common users and simplifies permission management operations for administrators, improving usability and user experience of tenant permission management.

- FusionInsight Manager employs role-based access control (RBAC) to configure different permissions for users based on service scenarios during multi-tenant management.
- The administrator of tenants has tenant management permissions, including viewing resources and services of the current tenant, adding or deleting sub-tenants of the current tenant, and managing permissions of sub-tenants' resources. FusionInsight Manager supports setting of the administrator for a single tenant so that the management over this tenant can be delegated to a user who is not the system administrator.

- Roles of a tenant have all permissions on the computing resources and storage resources of the tenant. When a tenant is created, the system automatically creates roles for this tenant. You can add a user and bind the user to the tenant roles so that the user can use the resources of the tenant.

Clear Resource Management

- **Self-Service Resource Configuration**

  In FusionInsight Manager, you can configure the computing resources and storage resources during the creation of a tenant and add, modify, or delete the resources of the tenant.

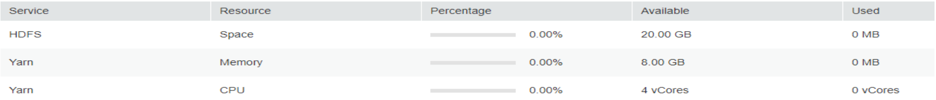
  Permissions of the roles that are associated with a tenant are updated automatically when you modify the computing or storage resources of the tenant.

- **Resource Usage Statistics**

  Resource usage statistics are critical for administrators to determine O&M activities based on the status of cluster applications and services, improving the cluster O&M efficiency. FusionInsight Manager displays the resource statistics of tenants in **Resource Quota**, including the vCores, memory, and HDFS storage resources.

  NOTE

  - **Resource Quota** dynamically calculates the resource usage of tenants.

Service	Resource	Percentage	Available	Used
HDFS	Space	0.00%	20.00 GB	0 MB
Yarn	Memory	0.00%	8.00 GB	0 MB
Yarn	CPU	0.00%	4 vCores	0 vCores

    The available resources of the Superior scheduler are calculated as follows:

    - Superior

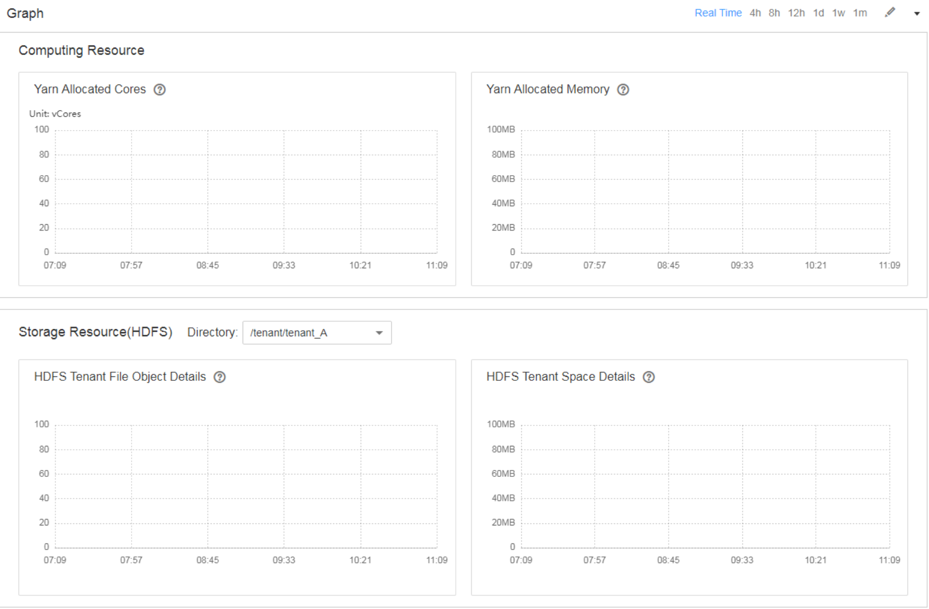
      The available Yarn resources (memory and CPU) are allocated in proportion based on the queue weight.

  - When the tenant administrator is bound to a tenant role, the tenant administrator has the permissions to manage the tenant and use all resources of the tenant.

- **Graphical Resource Monitoring**

  Graphical resource monitoring supports the graphical display of monitoring metrics listed in **Table 9-29**, as shown in **Figure 9-9**.

Figure 9-9 Refined monitoring



By default, the real-time monitoring data is displayed. You can click ✎ to customize a time range. The default time ranges include 4 hours, 8 hours, 12 hours, 1 day, 1 week, and 1 month. Click ▼ and select **Export** to export the monitoring metric information.

Table 9-29 Monitoring metrics

Service	Metric Item	Description
HDFS	HDFS Tenant Space Details ● Allocated Space ● Used Space	HDFS can monitor a specified storage directory. The storage directory is the same as the directory added by the current tenant in **Resource**.
	HDFS Tenant File Object Details ● Number of Used File Objects	
Yarn	Yarn Allocated Cores ● Maximum Number of CPU Cores in an AM ● Allocated Cores ● Number of Used CPU Cores in an AM	Monitoring information of the current tenant is displayed. If no sub-item is configured for a tenant, this information is not displayed. The monitoring data is obtained from **Scheduler** > **Application Queues** > **Queue:** *Tenant name* on the native web UI of Yarn.

Service	Metric Item	Description
	Yarn Allocated Memory ● Allocated Maximum AM Memory ● Allocated Memory ● Used AM Memory	

9.7.1.2.2 Multi-Tenant Model

Related Model

The following figure shows a multi-tenant model.

Figure 9-10 Multi-tenant model

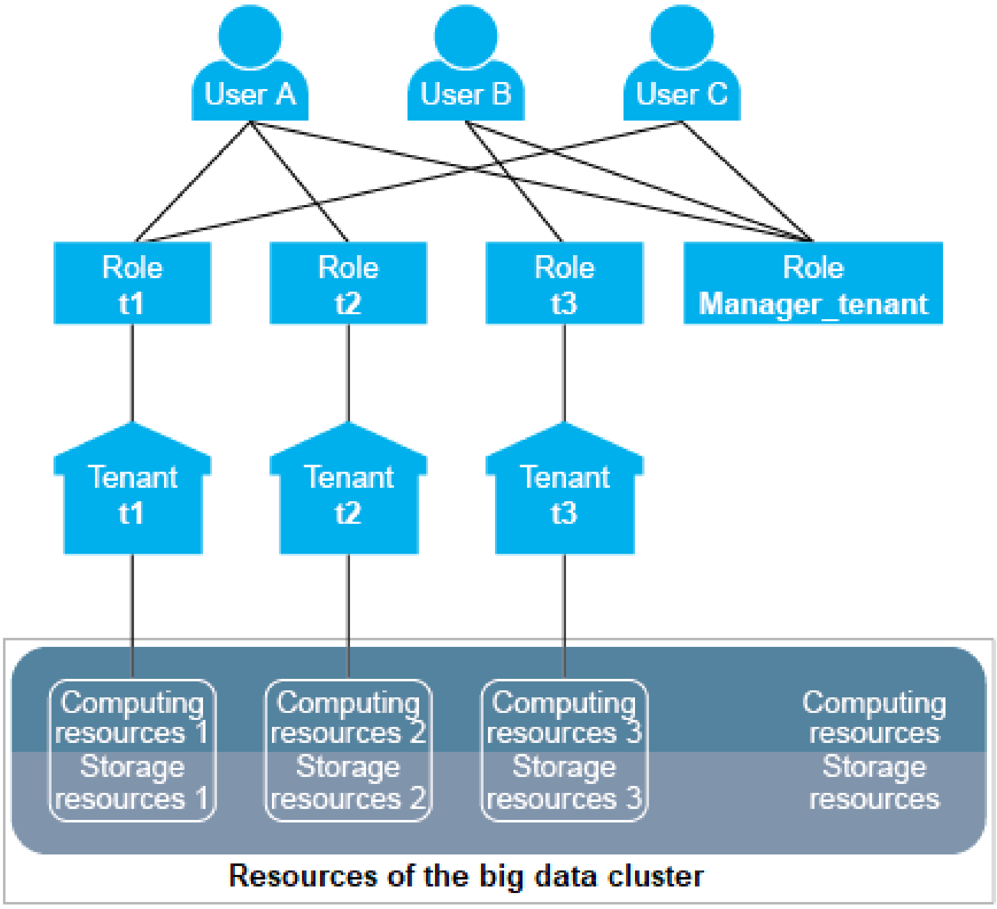


Table 9-30 describes the concepts involved in **Figure 9-10**.

Table 9-30 Concepts in the model

Concept	Description
User	A natural person who has a username and password and uses the big data cluster. There are three different users in **Figure 9-10**: user A, user B, and user C.
Role	A role is a carrier of one or more permissions. Permissions are assigned to specific objects, for example, access permissions for the **/tenant** directory in HDFS. **Figure 9-10** shows four roles: **t1**, **t2**, **t3**, and **Manager_tenant**. ● Roles **t1**, **t2**, and **t3** are automatically generated when tenants are created. The role names are the same as the tenant names. That is, roles **t1**, **t2**, and **t3** map to tenants **t1**, **t2**, and **t3**. Role names and tenant names need to be used in pair. ● Role **Manager_tenant** is defaulted in the cluster and cannot be used separately.
Tenant	A tenant is a resource set in a big data cluster. Multiple tenants are referred to as multi-tenancy. The resource sets further divided under a tenant are called sub-tenants. **Figure 9-10** shows three tenants: **t1**, **t2**, and **t3**.
Resource	● Computing resources include CPUs and memory. The computing resources of a tenant are allocated from the total computing resources in the cluster. One tenant cannot occupy the computing resources of another tenant. In **Figure 9-10**, computing resources 1, 2, and 3 are allocated for tenants **t1**, **t2**, and **t3** respectively from the cluster's computing resources. ● Storage resources include disks and third-party storage systems. The storage resources of a tenant are allocated from the total storage resources in the cluster. One tenant cannot occupy the storage resources of another tenant. In **Figure 9-10**, storage resources 1, 2, and 3 are allocated for tenants **t1**, **t2**, and **t3** respectively from the cluster's storage resources.

If a user wants to use a tenant's resources or add or delete a sub-tenant of a tenant, the user needs to be bound to both the tenant role and role **Manager_tenant**. **Table 9-31** lists the roles bound to each user in **Figure 9-10**.

Table 9-31 Roles bound to each user

User	Role	Permission
User A	- Role **t1** - Role **t2** - Role **Manager_tenant**	- Uses the resources of tenants **t1** and **t2**. - Adds or deletes sub-tenants of tenants **t1** and **t2**.
User B	- Role **t3** - Role **Manager_tenant**	- Uses the resources of tenant **t3**. - Adds or deletes sub-tenants of tenant **t3**.
User C	- Role **t1** - Role **Manager_tenant**	- Uses the resources of tenant **t1**. - Adds or deletes sub-tenants of tenant **t1**.

A user can be bound to multiple roles, and one role can also be bound to multiple users. Users are associated with tenants after being bound to the tenant roles. Therefore, tenants and users form a many-to-many relationship. One user can use the resources of multiple tenants, and multiple users can use the resources of the same tenant. For example, in **Figure 9-10**, user A uses the resources of tenants **t1** and **t2**, and users A and C uses the resources of tenant **t1**.

NOTE

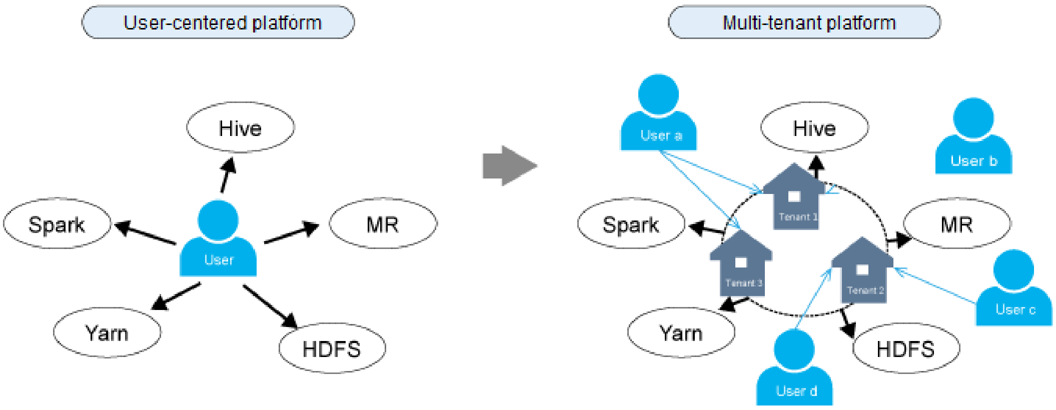
The concepts of a parent tenant, sub-tenant, level-1 tenant, and level-2 tenant are all designed for the multi-tenant service scenarios. Pay attention to the differences these concepts and the concepts of a leaf tenant resource and non-leaf tenant resource on FusionInsight Manager.

- Level-1 tenant: determined based on the tenant's level. For example, the first created tenant is a level-1 tenant and its sub-tenant is a level-2 tenant.
- Parent tenant and sub-tenant: indicates the hierarchical relationship between tenants.
- Non-leaf tenant resource: indicates the tenant type selected during tenant creation. This tenant type can be used to create sub-tenants.
- Leaf tenant resource: indicates the tenant type selected during tenant creation. This tenant type cannot be used to create sub-tenants.

Multi-Tenant Platform

Tenant is a core concept of the FusionInsight big data platform. It plays an important role in big data platforms' transformation from user-centered to multi-tenant to keep up with enterprises' multi-tenant application environments. **Figure 9-11** shows the transformation of big data platforms.

Figure 9-11 Platform transformation from user-centered to multi-tenant



On a user-centered big data platform, users can directly access and use all resources and services.

- However, user applications may use only partial cluster resources, resulting in low resource utilization.
- The data of different users may be stored together, decreasing data security.

On a multi-tenant big data platform, users use required resources and services by accessing the tenants.

- Resources are allocated and scheduled based on application requirements and used based on tenants, increasing resource utilization.
- Users can access the resources of tenants only after being associated with tenant roles, enhancing access security.
- The data of tenants is isolated, ensuring data security.

9.7.1.2.3 Resource Overview

MRS cluster resources are classified into computing resources and storage resources. The multi-tenant architecture implements resource isolation.

- **Computing resources**

  Computing resources include CPUs and memory. One tenant cannot occupy the computing resources of another tenant.
- **Storage resources**

  Storage resources include disks and third-party storage systems. One tenant cannot access the data of another tenant.

Computing Resources

Computing resources are divided into static service resources and dynamic resources.

- **Static Service Resources**

  Static service resources are computing resources allocated to each service and are not shared between services. The total computing resources of each service are fixed. These services include Flume, HBase, HDFS, and Yarn.
- **Dynamic Resources**

Dynamic resources are computing resources dynamically scheduled to a job queue by the distributed resource management service Yarn. Yarn dynamically schedules resources for the job queues of MapReduce, Spark2x, Flink, and Hive.

NOTE

The resources allocated to Yarn in a big data cluster are static service resources but can be dynamically allocated to job queues by Yarn.

Storage Resources

Storage resources are data storage resources that can be allocated by the distributed file storage service HDFS. Directory is the basic unit of allocating HDFS storage resources. Tenants can obtain storage resources from the specified directories in the HDFS file system.

9.7.1.2.4 Dynamic Resources

Overview

Yarn provides distributed resource management for a big data cluster. The total volume of resources allocated to Yarn can be configured. Then Yarn allocates and schedules computing resources for job queues. The computing resources of MapReduce, Spark, Flink, and Hive job queues are allocated and scheduled by Yarn.

Yarn queues are fundamental units of scheduling computing resources.

The resources obtained by tenants using Yarn queues are dynamic resources. Users can dynamically create and modify the queue quotas and view the status and statistics of the queues.

Resource Pools

Nowadays, enterprise IT systems often face complex cluster environments and diverse upper-layer requirements. For example:

- Heterogeneous cluster: The computing speed, storage capacity, and network performance of each node in the cluster are different. All the tasks of complex applications need to be properly allocated to each compute node in the cluster based on service requirements.
- Computing isolation: Data must be shared among multiple departments but computing resources must be distributed onto different compute nodes.

These require that the compute nodes be further partitioned.

Resource pools are used to specify the configuration of dynamic resources. Yarn queues are associated with resource pools for resource allocation and scheduling.

One tenant can have only one default resource pool. Users can be bound to the role of a tenant to use the resources in the resource pool of the tenant. To use resources in multiple resource pools, a user can be bound to roles of multiple tenants.

Scheduling Mechanism

Yarn dynamic resources support label-based scheduling. This policy creates labels for compute nodes (Yarn NodeManagers) and adds the compute nodes with the same label into the same resource pool. Then Yarn dynamically associates the queues with resource pools based on the resource requirements of the queues.

For example, a cluster has more than 40 nodes which are labeled by **Normal**, **HighCPU**, **HighMEM**, or **HighIO** based on their hardware and network configurations and added into four resource pools, respectively. **Table 9-32** describes the performance of each node in the resource pool.

Table 9-32 Performance of each node in a resource pool

Label	Number of Nodes	Hardware and Network Configuration	Added To	Associated With
Normal	10	General	Resource pool A	Common queue
HighCPU	10	High-performance CPU	Resource pool B	Computing-intensive queue
HighMEM	10	Large memory	Resource pool C	Memory-intensive queue
HighIO	10	High-performance network	Resource pool D	I/O-intensive queue

A queue can use only the compute nodes in its associated resource pool.

- A common queue is associated with resource pool A and uses **Normal** nodes with general hardware and network configurations.
- A computing-intensive queue is associated with resource pool B and uses **HighCPU** nodes with high-performance CPUs.
- A memory-intensive queue is associated with resource pool C and uses **HighMEM** nodes with large memory.
- An I/O-intensive queue is associated with resource pool C and uses **HighIO** nodes with high-performance network.

Yarn queues are associated with specified resource pools to efficiently utilize resources in resource pools and maximize node performance.

FusionInsight Manager supports a maximum of 50 resource pools. The system has a default resource pool.

Schedulers

By default, the Superior scheduler is enabled for the MRS cluster.

- The Superior scheduler is an enhanced version and named after the Lake Superior, indicating that the scheduler can manage a large amount of data.

To meet enterprise requirements and tackle scheduling challenges faced by the Yarn community, the Superior scheduler makes the following enhancements:

- Enhanced resource sharing policy

  The Superior scheduler supports queue hierarchy. It integrates the functions of open-source schedulers and shares resources based on configurable policies. In terms of instances, administrators can use the Superior scheduler to configure an absolute value or percentage policy for queue resources. The resource sharing policy of the Superior scheduler enhances label-based scheduling of Yarn as a resource pool feature. The nodes in the Yarn cluster can be grouped based on the capacity or service type to ensure that queues can more efficiently utilize resources.
- Tenant-based resource reservation policy

  Some tenants may run critical tasks at some time, and their resource requirements must be preferentially addressed. The Superior scheduler builds a mechanism to support the resource reservation policy. Reserved resources can be allocated to the critical tasks running in the specified tenant queues in a timely manner to ensure proper task execution.
- Fair sharing among tenants and resource pool users

  The Superior scheduler allows shared resources to be configured for users in a queue. Each tenant may have users with different weights. Heavily weighted users may require more shared resources.
- Ensured scheduling performance in a big cluster

  The Superior scheduler receives heartbeats from each NodeManager and saves resource information in memory, which enables the scheduler to control cluster resource usage globally. The Superior scheduler uses the push scheduling model, which makes the scheduling more precise and efficient and remarkably improves cluster resource utilization. Additionally, the Superior scheduler delivers excellent performance when the interval between NodeManager heartbeats is long and prevents heartbeat storms in big clusters.
- Priority policy

  If the minimum resource requirement of a service cannot be met after the service obtains all available resources, a preemption occurs. The preemption function is disabled by default.

9.7.1.2.5 Storage Resources

Overview

As a distributed file storage service in a big data cluster, HDFS stores all the user data of the upper-layer applications in the big data cluster, including the data written to HBase tables or Hive tables.

A directory is the basic unit of allocating HDFS storage resources. HDFS supports the conventional hierarchical file structure. Users or applications can create directories and create, delete, move, or rename files in directories. Tenants can obtain storage resources from specified directories in the HDFS file system.

Scheduling Mechanism

HDFS directories can be stored on nodes with specified labels or disks of specified hardware types. For example:

- When both real-time query and data analysis tasks are running in the same cluster, the real-time query tasks need to be deployed only on certain nodes, and the task data must also be stored on these nodes.
- Based on actual service requirements, key data needs to be stored on highly reliable nodes.

Administrators can flexibly configure HDFS data storage policies based on actual service requirements and data features to store data on specified nodes.

For tenants, storage resources refer to the HDFS resources they use. Data of specified directories can be stored to the tenant-specified storage paths, thereby implementing storage resource scheduling and ensuring data isolation between tenants.

Users can add or delete HDFS storage directories of tenants and set the file quantity quota and storage capacity quota of directories to manage storage resources.

9.7.1.3 Multi-Tenancy Usage

9.7.1.3.1 Overview

Tenants are used in resource control and service isolation scenarios. Administrators need to determine the service scenarios of cluster resources and then plan tenants.

NOTE

- Yarn in a new cluster uses the Superior scheduler by default. For details, see **Using the Superior Scheduler**.

Multi-tenancy involves three types of operations: creating a tenant, managing tenants, and managing resources. **Table 9-33** describes these operations.

Table 9-33 Multi-tenant operations

Operation	Action	Description
Creating a tenant	● Add a tenant. ● Add a sub-tenant. ● Create a user and bind the user to the role of a tenant.	During the creation of a tenant, you can configure its computing resources, storage resources, and associated services based on service requirements. In addition, you can add users to the tenant and bind necessary roles to these users. A user to create a level-1 tenant needs to be bound to the **Manager_administrator** or **System_administrator** role. A user to create a sub-tenant needs to be bound to the role of the parent tenant at least.

Operation	Action	Description
Managing tenants	● Manage the tenant directory. ● Restore tenant data. ● Clear non-associated queues of a tenant. ● Delete a tenant.	You can edit tenants as services change. A user to manage or delete a level-1 tenant or restore tenant data needs to be bound to the **Manager_administrator** or **System_administrator** role. A user to manage or delete a sub-tenant needs to be bound to the role of the parent tenant at least.
Managing resources	● Create a resource pool. ● Modify a resource pool. ● Delete a resource pool. ● Configure a queue. ● Configure the queue capacity policy of a resource pool. ● Clear configurations of a queue.	You can reconfigure resources for tenants as the services change. A user to manage resources needs to be bound to the **Manager_administrator** or **System_administrator** role.

9.7.1.3.2 Process Overview

Administrators need to determine the service scenarios of cluster resources and then plan tenants. After that, administrators add tenants and configure dynamic resources, storage resources, and associated services for the tenants on FusionInsight Manager.

Process Overview shows the process for creating a tenant.

Figure 9-12 Creating a tenant

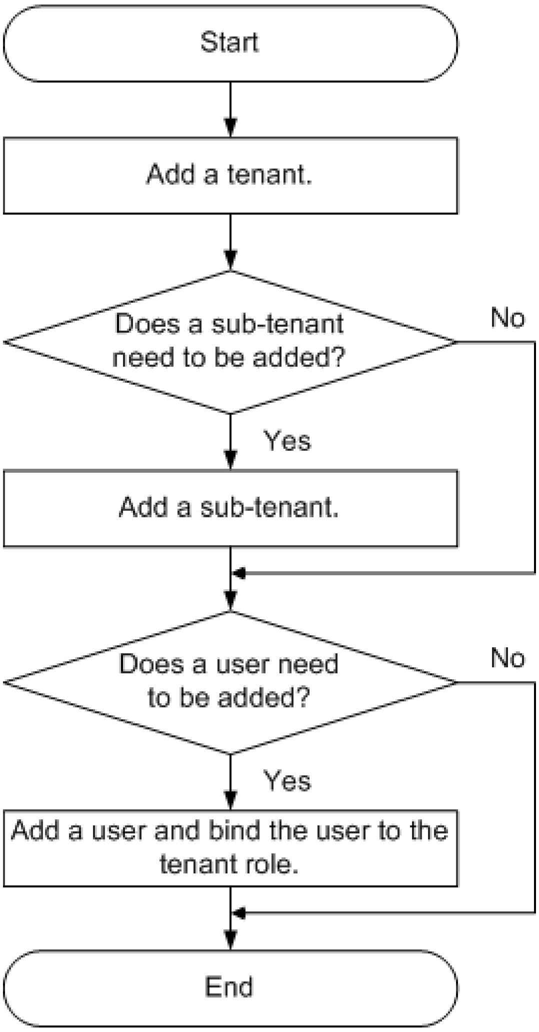

Start
Add a tenant.
Does a sub-tenant need to be added?
No
Yes
Add a sub-tenant.
Does a user need to be added?
No
Yes
Add a user and bind the user to the tenant role.
End

Table 9-34 describes the operations for creating a tenant.

Table 9-34 Operations for creating a tenant

Operation	Description
Add a tenant.	You can configure the computing resources, storage resources, and associated services of the tenant.
Add a sub-tenant.	You can configure the computing resources, storage resources, and associated services of the sub-tenant.
Add a user and bind the user to the tenant role.	If a user wants to use the resources of tenant **tenant1** or add or delete sub-tenants for **tenant1**, the user must be bound to both the **Manager_tenant** and **tenant1**_*Cluster ID* roles.

9.7.2 Using the Superior Scheduler

9.7.2.1 Creating Tenants

9.7.2.1.1 Adding a Tenant

Scenario

You can create tenants on FusionInsight Manager based on the resource consumption and isolation planning and requirements of services.

Prerequisites

- A tenant name has been planned based on service requirements. The name cannot be the same as that of a role, HDFS directory, or Yarn queue that exists in the current cluster.
- Resources to be allocated to the current tenant have been planned to ensure that the sum of resources of direct sub-tenants at each level does not exceed the resources of the current tenant.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Tenant Resources**.

Step 2 Click ⊕. On the page that is displayed, configure tenant attributes according to **Table 9-35**.

Table 9-35 Tenant parameters

Parameter	Description
Cluster	Indicates the cluster for which you want to create a tenant.
Name	● Indicates the name of the current tenant. The value consists of 3 to 50 characters, including digits, letters, and underscores (_). ● Plan a tenant name based on service requirements. The name cannot be the same as that of a role, HDFS directory, or Yarn queue that exists in the current cluster.
Tenant Resource Type	Specifies whether the tenant is a leaf tenant. ● When **Leaf Tenant Resource** is selected, the current tenant is a leaf tenant and no sub-tenant can be added. ● When **Non-leaf Tenant Resource** is selected, the current tenant is not a leaf tenant and sub-tenants can be added to the current tenant.

Parameter	Description
Computing Resource	Specifies the dynamic computing resources for the current tenant. ● When **Yarn** is selected, the system automatically creates a queue in Yarn and the queue is named the same as the tenant name. – A leaf tenant can directly submit jobs to the queue. – A non-leaf tenant cannot directly submit jobs to the queue. However, Yarn adds an extra queue (hidden) named **default** for the non-leaf tenant to record the remaining resource capacity of the tenant. Actual jobs do not run in this queue. ● If **Yarn** is not selected, the system does not automatically create a queue.
Configuration Mode	Indicates the configuration mode of computing resource parameters. ● If you select **Basic**, you only need to set **Default Resource Pool Capacity (%)**. ● If you select **Advanced**, you can manually configure the resource allocation weight and the minimum, maximum, and reserved resources of the tenant.
Default Resource Pool Capacity (%)	Indicates the percentage of computing resources used by the current tenant in the default resource pool. The value ranges from **0** to **100%**.
Weight	Indicates the resource allocation weight. The value ranges from **0** to **100**.
Minimum Resource	Indicates the resources guaranteed for the tenant (preemption supported). The value can be a percentage or an absolute value of the parent tenant's resources. When a tenant has a light workload, the resources of the tenant are automatically allocated to other tenants. When the available tenant resources are less than the value of **Minimum Resource**, the tenant can preempt the resources that have been lent to other tenants.
Maximum Resource	Indicates the maximum resources that can be used by the tenant. The tenant cannot obtain more resources than the value configured. The value can be a percentage or an absolute value of the parent tenant's resources.

Parameter	Description
Reserved Resource	Indicates the resources reserved for the tenant. The reserved resources cannot be used by other tenants even if no job is running in the current tenant resources. The value can be a percentage or an absolute value of the parent tenant's resources.
Storage Resource	Specifies storage resources for the current tenant. ● When **HDFS** is selected, the system automatically allocates storage resources. ● When **HDFS** is not selected, the system does not automatically allocate storage resources.
Quota	Indicates the quota for files and directories.
Space Quota	Indicates the quota for the HDFS storage space used by the current tenant. ● If the unit is set to **MB**, the value ranges from **1** to **8796093022208**. If the unit is set to **GB**, the value ranges from **1** to **8589934592**. ● This parameter indicates the maximum HDFS storage space that can be used by the tenant, but not the actual space used. ● If its value is greater than the size of the HDFS physical disk, the maximum space available is the full space of the HDFS physical disk.
Storage Path	Indicates the HDFS storage directory for the tenant. ● The system automatically creates a folder named after the tenant name in the **/tenant** directory by default. For example, the default HDFS storage directory for tenant **ta1** is **/tenant/ta1**. ● When a tenant is created for the first time, the system creates the **/tenant** directory in the HDFS root directory. The storage path is customizable.
Service	Specifies whether to associate resources of other services. For details, see **Step 4**.
Description	Indicates the description of the current tenant.

NOTE

Roles, computing resources, and storage resources are automatically created when tenants are created.

- The new role has permissions on the computing and storage resources. This role and its permissions are automatically controlled by the system and cannot be manually managed by choosing **System** > **Permission** > **Role**. The role name is in the format of *Tenant name_Cluster ID*. The ID of the first cluster is not displayed by default.
- When using this tenant, create a system user and bind the user to the role of the tenant. For details, see **Adding a User and Binding the User to a Tenant Role**.
- During the tenant creation, the system automatically creates a Yarn queue named after the tenant. If the queue name already exists, the new queue is named **Tenant name-***N*. *N* indicates a natural number starting from **1**. When a same name exists, the value *N* increases automatically to differentiate the queue from others. For example, **saletenant**, **saletenant-1**, and **saletenant-2**.

Step 3 Check whether the current tenant needs to be associated with resources of other services.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 Click **Associate Service** to configure other service resources used by the current tenant, and click **OK**.

- Set **Service** to **HBase** and **Association Type** to **Exclusive** or **Shared**.

  NOTE

  - **Exclusive** indicates that the service resources are used by the tenant exclusively and cannot be associated with other tenants.
  - **Shared** indicates that the service resources can be shared with other tenants.

NOTE

- Only HBase can be associated with a new tenant. However, HDFS, HBase, and Yarn can be associated with existing tenants.
- To associate an existing tenant with service resources, click the target tenant in the tenant list, switch to the **Service Associations** page, and click **Associate Service** to configure resources to be associated with the tenant.
- To disassociate an existing tenant from service resources, click the target tenant in the tenant list, switch to the **Service Associations** page, and click **Delete** in the **Operation** column. In the displayed dialog box, select **I have read the information and understand the impact** and click **OK**.

Step 5 Click **OK**. Wait until the system displays a message indicating that the tenant is successfully created.

----End

9.7.2.1.2 Adding a Sub-Tenant

Scenario

You can create sub-tenants on FusionInsight Manager and allocate resources of the current tenant to the sub-tenants based on the resource consumption and isolation planning and requirements of services.

Prerequisites

- A parent non-leaf tenant has been added.
- A sub-tenant name has been planned based on service requirements. The name cannot be the same as that of a role, HDFS directory, or Yarn queue that exists in the current cluster.
- Resources to be allocated to the current tenant have been planned to ensure that the sum of resources of direct sub-tenants at each level does not exceed the resources of the current tenant.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Tenant Resources**.

Step 2 In the tenant list on the left, select a parent tenant and click ⊕. On the page for adding a sub-tenant, set attributes for the sub-tenant according to **Table 9-36**.

Table 9-36 Sub-tenant parameters

Parameter	Description
Cluster	Indicates the cluster to which the parent tenant belongs.
Parent Tenant Resource	Indicates the name of the parent tenant.
Name	● Indicates the name of the current tenant. The value consists of 3 to 50 characters, including digits, letters, and underscores (_). ● Plan a sub-tenant name based on service requirements. The name cannot be the same as that of a role, HDFS directory, or Yarn queue that exists in the current cluster.
Tenant Resource Type	Specifies whether the tenant is a leaf tenant. ● When **Leaf Tenant Resource** is selected, the current tenant is a leaf tenant and no sub-tenant can be added. ● When **Non-leaf Tenant Resource** is selected, the current tenant is not a leaf tenant and sub-tenants can be added to the current tenant. However, the tenant depth cannot exceed 5 levels.

Parameter	Description
Computing Resource	Specifies the dynamic computing resources for the current tenant. ● When **Yarn** is selected, the system automatically creates a queue in Yarn and the queue is named the same as the sub-tenant name. – A leaf tenant can directly submit jobs to the queue. – A non-leaf tenant cannot directly submit jobs to the queue. However, Yarn adds an extra queue (hidden) named **default** for the non-leaf tenant to record the remaining resource capacity of the tenant. Actual jobs do not run in this queue. ● If **Yarn** is not selected, the system does not automatically create a queue.
Configuration Mode	Indicates the configuration mode of computing resource parameters. ● If you select **Basic**, you only need to set **Default Resource Pool Capacity (%)**. ● If you select **Advanced**, you can manually configure the resource allocation weight and the minimum, maximum, and reserved resources of the tenant.
Default Resource Pool Capacity (%)	Indicates the percentage of computing resources used by the current tenant. The base value is the total resources of the parent tenant.
Weight	Indicates the resource allocation weight. The value ranges from **0** to **100**.
Minimum Resource	Indicates the resources guaranteed for the tenant (preemption supported). The value can be a percentage or an absolute value of the parent tenant's resources. When a tenant has a light workload, the resources of the tenant are automatically allocated to other tenants. When the available tenant resources are less than the value of **Minimum Resource**, the tenant can preempt the resources that have been lent to other tenants.
Maximum Resource	Indicates the maximum resources that can be used by the tenant. The tenant cannot obtain more resources than the value configured. The value can be a percentage or an absolute value of the parent tenant's resources.

Parameter	Description
Reserved Resource	Indicates the resources reserved for the tenant. The reserved resources cannot be used by other tenants even if no job is running in the current tenant resources. The value can be a percentage or an absolute value of the parent tenant's resources.
Storage Resource	Specifies storage resources for the current tenant. ● When **HDFS** is selected, the system automatically creates a folder named after the sub-tenant in the HDFS parent tenant directory. ● When **HDFS** is not selected, the system does not automatically allocate storage resources.
Quota	Indicates the quota for files and directories.
Space Quota	Indicates the quota for the HDFS storage space used by the current tenant. ● If the unit is set to **MB**, the value ranges from **1** to **8796093022208**. If the unit is set to **GB**, the value ranges from **1** to **8589934592**. ● This parameter indicates the maximum HDFS storage space that can be used by the tenant, but not the actual space used. ● If its value is greater than the size of the HDFS physical disk, the maximum space available is the full space of the HDFS physical disk. ● If this quota is greater than the quota of the parent tenant, the actual storage space does not exceed the quota of the parent tenant.
Storage Path	Indicates the HDFS storage directory for the tenant. ● The system automatically creates a folder named after the sub-tenant name in the directory of the parent tenant by default. For example, if the sub-tenant is **ta1s** and the parent directory is **/tenant/ta1**, the storage path for the sub-tenant is then **/tenant/ta1/ta1s**. ● The storage path is customizable in the parent directory.
Service	Specifies whether to associate resources of other services. For details, see **Step 4**.
Description	Indicates the description of the current tenant.

NOTE

Roles, computing resources, and storage resources are automatically created when tenants are created.

- The new role has permissions on the computing and storage resources. This role and its permissions are automatically controlled by the system and cannot be manually managed by choosing **System** > **Permission** > **Role**. The role name is in the format of *Tenant name_Cluster ID*. The ID of the first cluster is not displayed by default.
- When using this tenant, create a system user and bind the user to the role of the tenant. For details, see **Adding a User and Binding the User to a Tenant Role**.
- The sub-tenant can further allocate the resources of its parent tenant. The sum of the resource percentages of direct sub-tenants under a parent tenant at each level cannot exceed 100%. The sum of the computing resource percentages of all level-1 tenants cannot exceed 100%.

Step 3 Check whether the current tenant needs to be associated with resources of other services.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 Click **Associate Service** to configure other service resources used by the current tenant.

1. Set **Services** to **HBase**.
2. Set **Association Type** as follows:
   - **Exclusive** indicates that the service resources are used by the tenant exclusively and cannot be associated with other tenants.
   - **Shared** indicates that the service resources can be shared with other tenants.

   NOTE

   - Only HBase can be associated with a new tenant. However, HDFS, HBase, and Yarn can be associated with existing tenants.
   - To associate an existing tenant with service resources, click the target tenant in the tenant list, switch to the **Service Associations** page, and click **Associate Service** to configure resources to be associated with the tenant.
   - To disassociate an existing tenant from service resources, click the target tenant in the tenant list, switch to the **Service Associations** page, and click **Delete** in the **Operation** column. In the displayed dialog box, select **I have read the information and understand the impact** and click **OK**.
3. Click **OK**.

Step 5 Click **OK**. Wait until the system displays a message indicating that the tenant is successfully created.

----End

9.7.2.1.3 Adding a User and Binding the User to a Tenant Role

Scenario

A newly created tenant cannot directly log in to the cluster to access resources. You need to add a user for the tenant on FusionInsight Manager and bind the user to the role of the tenant to assign operation permissions to the user.

Prerequisites

You have clarified service requirements and created a tenant.

Procedure

Step 1 Log in to FusionInsight Manager and choose **System** > **Permission** > **User**.

Step 2 If you want to add a user to the system, click **Create**.

If you want to bind tenant roles to an existing user in the system, locate the row of the user and click **Modify** in the **Operation** column.

Set user attributes according to **Table 9-37**.

Table 9-37 User parameters

Parameter	Description
Username	Indicates the current username. The value contains 3 to 32 characters, including digits, letters, underscores (_), hyphens (-), and spaces. ● The username cannot be the same as the OS username of any node in the cluster. Otherwise, the user cannot be used. ● A username that differs only in alphabetic case from an existing username is not allowed. For example, if **User1** has been created, you cannot create **user1**. Enter the correct username when using **User1**.
User Type	The options are **Human-Machine** and **Machine-Machine**. ● **Human-Machine** user: used for FusionInsight Manager O&M and component client operations. If you select this option, set both **Password** and **Confirm Password** accordingly. ● **Machine-Machine** user: used for application development. If you select this option, the password is randomly generated.
Password	This parameter is mandatory if **User Type** is set to **Human-Machine**. The password must contain 8 to 64 characters of at least four types of the following: uppercase letters, lowercase letters, digits, special characters, and spaces. The password cannot be the username or the username spelled backwards.
Confirm Password	Enter the password again.

Parameter	Description
User Group	In the **User Group** area, click **Add** and select user groups to add the user to the groups. ● If roles have been added to the user groups, the user can be granted the permissions of the roles. ● For example, add the user to the Hive user group to assign Hive permissions to the user.
Primary Group	Select a group as the primary group for the user to create directories and files. The drop-down list contains all groups selected in **User Group**.
Role	Click **Add** to bind a tenant role to the user. **NOTE** ● If a user wants to use the resources of tenant **tenant1** and to add or delete sub-tenants for **tenant1**, the user must be bound to both the **Manager_tenant** and **tenant1**_*Cluster ID* roles. ● If the tenant has been associated with the HBase service and Ranger authentication is enabled for the cluster, you need to configure the HBase execution permissions on the Ranger page.
Description	Indicates the description of the current user.

Step 3 Click **OK**.

----End

9.7.2.2 Managing Tenants

9.7.2.2.1 Managing Tenant Directories

Scenario

You can manage the HDFS storage directories used by specified tenants based on service requirements on FusionInsight Manager, such as adding tenant directories, changing the quotas for directories and files and for storage space, and deleting directories.

Prerequisites

A tenant with HDFS storage resources has been added.

Viewing a Tenant Directory

Step 1 Log in to FusionInsight Manager and choose **Tenant Resources**.

Step 2 In the tenant list on the left, click the target tenant.

Step 3 Click the **Resource** tab.

Step 4 View the **HDFS Storage** table.

- The **File Number Threshold** column provides the quota for files and directories of the tenant directory.
- The **Space Quota** column provides the storage space size of the tenant directory.

----End

Adding a Tenant Directory

Step 1 On FusionInsight Manager, choose **Tenant Resources**.

Step 2 In the tenant list on the left, click the target tenant.

Step 3 Click the **Resource** tab.

Step 4 In the **HDFS Storage** area, click **Create Directory**.

- **Parent Directory**: indicates the storage directory used by the parent tenant of the current tenant.

  NOTE

  This parameter is not displayed if the current tenant is not a sub-tenant.

- Set **Path** to a tenant directory path.

  NOTE

  If the current tenant is not a sub-tenant, the new path is created in the HDFS root directory.

- Set **Quota** to the quota for files and directories.
- **File Number Threshold (%)** is valid only when **Quota** is set. If the ratio of the number of used files to the value of **Quota** exceeds the value of this parameter, an alarm is generated. If this parameter is not specified, no alarm is reported in this scenario.

  NOTE

  The number of used files is collected every hour. Therefore, the alarm indicating that the ratio of used files exceeds the threshold is delayed.

- Set **Space Quota** to the storage space size of the tenant directory.
- If the ratio of used storage space to the value of **Space Quota** exceeds the **Storage Space Threshold (%)** value, an alarm is generated. If this parameter is not specified, no alarm is reported in this scenario.

  NOTE

  The used storage space is collected every hour. Therefore, the alarm indicating that the ratio of used storage space exceeds the threshold is delayed.

Step 5 Click **OK**.

----End

Modifying a Tenant Directory

Step 1 On FusionInsight Manager, choose **Tenant Resources**.

Step 2 In the tenant list on the left, click the target tenant.

Step 3 Click the **Resource** tab.

Step 4 In the **HDFS Storage** table, click **Modify** in the **Operation** column of the specified tenant directory.

- Set **Quota** to the quota for files and directories.
- **File Number Threshold (%)** is valid only when **Quota** is set. If the ratio of the number of used files to the value of **Quota** exceeds the value of this parameter, an alarm is generated. If this parameter is not specified, no alarm is reported in this scenario.
- Set **Space Quota** to the storage space size of the tenant directory.
- If the ratio of used storage space to the value of **Space Quota** exceeds the **Storage Space Threshold (%)** value, an alarm is generated. If this parameter is not specified, no alarm is reported in this scenario.

Step 5 Click **OK**.

----End

Deleting a Tenant Directory

Step 1 On FusionInsight Manager, choose **Tenant Resources**.

Step 2 In the tenant list on the left, click the target tenant.

Step 3 Click the **Resource** tab.

Step 4 In the **HDFS Storage** table, click **Delete** in the **Operation** column of the specified tenant directory.

NOTE

The tenant directory that is created by the system during tenant creation cannot be deleted.

Step 5 Click **OK**.

----End

9.7.2.2.2 Restoring Tenant Data

Scenario

Tenant data is stored on FusionInsight Manager and cluster components. When components are recovered from failures or reinstalled, some configuration data of all tenants may become abnormal. In this case, you need to manually restore the configuration data on FusionInsight Manager.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Tenant Resources**.

Step 2 In the tenant list on the left, click the target tenant.

Step 3 Check the tenant data status.

1. On the **Summary** page, check **Tenant Status**. A green icon indicates that the tenant is available and gray indicates that the tenant is unavailable.
2. Click **Resource** and check the icons on the left of **Yarn** and **HDFS Storage**. A green icon indicates that the resource is available, and gray indicates that the resource is unavailable.
3. Click **Service Associations** and check the **Status** column of the associated services. **Normal** indicates that the component can provide services for the associated tenant. **Not Available** indicates that the component cannot provide services for the tenant.
4. If any of the preceding check items is abnormal, go to **Step 4** to restore tenant data.

Step 4 Click . In the displayed dialog box, enter the password of the current login user and click **OK**.

Step 5 In the **Restore Tenant Resource Data** window, select one or more components to restore data, and click **OK**. The system automatically restores the tenant data.

----End

9.7.2.2.3 Deleting a Tenant

Scenario

You can delete tenants that are no longer used on FusionInsight Manager based on service requirements to release resources occupied by the tenants.

Prerequisites

- A tenant has been added.
- The tenant has no sub-tenants. If the tenant has sub-tenants, delete them; otherwise, the tenant cannot be deleted.
- The role of the tenant is not associated with any user or user group.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Tenant Resources**.

Step 2 In the tenant list on the left, click the target tenant and click .

NOTE

- If you want to retain the tenant data, select **Reserve the data of this tenant resource**. Otherwise, the storage space of the tenant will be deleted.

Step 3 Click **OK**.

It takes a few minutes to save the configuration. After the tenant is deleted, the role and storage space of the tenant are also deleted.

NOTE

After the tenant is deleted, the queue of the tenant still exists in Yarn. The queue of the tenant is not displayed on the role management page in Yarn.

----End

9.7.2.3 Managing Resources

9.7.2.3.1 Adding a Resource Pool

Scenario

In a cluster, you can logically group Yarn NodeManagers into Yarn resource pools. Each NodeManager belongs to only one resource pool. You can create a custom resource pool on FusionInsight Manager and add the hosts that have not been added to any custom resource pools to this resource pool so that specified queues can use the computing resources provided by these hosts.

The system contains a **default** resource pool by default. All NodeManagers that are not added to custom resource pools belong to this resource pool.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Tenant Resources** > **Resource Pool**.

Step 3 Click **Add Resource Pool**.

Step 4 Set resource pool attributes.

- **Cluster**: Select the cluster to which the resource pool is to be added.
- **Name**: Enter the name of the resource pool. The name contains 1 to 50 characters, including digits, letters, and underscores (_), and cannot start with an underscore (_).
- **Resource Label**: Enter the resource label of the resource pool. The value can contain 1 to 50 characters, including digits, letters, underscores (_), and hyphens (-), and must start with a digit or letter.
- **Resource**: In the **Available Hosts** area, select specified hosts and click >> to add the hosts to the **Selected Hosts** area. Only hosts in the cluster can be selected. The host list in the resource pool can be left blank.

  NOTE

  You can filter hosts by host name, number of CPU cores, memory, operating system, or platform type based on service requirements.

Step 5 Click **OK**.

After the resource pool is created, you can view its name, members, and mode in the resource pool list. Hosts that are added to the custom resource pool are no longer members of the **default** resource pool.

----End

9.7.2.3.2 Modifying a Resource Pool

Scenario

When hosts in a resource pool need to be adjusted based on service requirements, you can modify members in the resource pool on FusionInsight Manager.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Tenant Resources** > **Resource Pool**.

Step 3 Locate the row that contains the specified resource pool, and click **Edit** in the **Operation** column.

Step 4 In the **Resource** area, modify hosts.

- Adding hosts: Select desired hosts in **Available Hosts** and click >> to add them to the resource pool.
- Deleting hosts: Select desired hosts in **Selected Hosts** and click << to remove them from the resource pool. The host list in the resource pool can be left blank.

Step 5 Click **OK**.

----End

9.7.2.3.3 Deleting a Resource Pool

Scenario

If a resource pool is no longer used based on service requirements, you can delete it on FusionInsight Manager.

Prerequisites

- Any queue in the cluster does not use the resource pool to be deleted as the default resource pool. Before deleting the resource pool, cancel the default resource pool. For details, see **Configuring a Queue**.
- Resource distribution policies of all queues have been cleared from the resource pool to be deleted. For details, see **Clearing Queue Configurations**.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Tenant Resources** > **Resource Pool**.

Step 3 Locate the row that contains the specified resource pool, and click **Delete** in the **Operation** column.

Step 4 In the displayed dialog box, click **OK**.

----End

9.7.2.3.4 Configuring a Queue

Scenario

You can modify the queue configurations for a specified tenant on FusionInsight Manager.

Prerequisites

A tenant who uses the Superior scheduler has been added.

Procedure

Step 1 On FusionInsight Manager, choose **Tenant Resources**.

Step 2 Choose **Dynamic Resource Plan**.

Step 3 Click the **Queue Configurations** tab.

Step 4 Set **Cluster** to the name of the target cluster. In **All tenants resources** area, locate the row that contains the target tenant resource and click **Modify** in the **Operation** column.

NOTE

- You can also access the **Modify Queue Configuration** page as follows: In the tenant list on the **Tenant Resources Management** page, click the target tenant, click the **Resource** tab, and click ✎ next to **Queue Configurations (*Queue name*)**.
- A queue can be bound to only one non-default resource pool.
- For parameters such as **Max Allocated vCores**, **Max Allocated Memory(MB)**, **Max Running Apps**, **Max Running Apps per User**, and **Max Pending Apps**, if the value of a sub-tenant is **-1**, the value of the parent tenant can be set to a specific limit. If the parent tenant value is a specific limit, the sub-tenant value can be set to **-1**.
- **Max Allocated vCores** and **Max Allocated Memory(MB)** must be both changed to values other than **-1**.

Table 9-38 Queue configuration parameters

Parameter	Description
Max Master Shares(%)	Indicates the maximum percentage of resources occupied by all ApplicationMasters in the current queue.
Max Allocated vCores	Indicates the maximum number of cores that can be allocated to a single Yarn container in the current queue. The default value is **-1**, indicating that the number of cores is not limited within the value range.
Max Allocated Memory(MB)	Indicates the maximum memory that can be allocated to a single Yarn container in the current queue. The default value is **-1**, indicating that the memory is not limited within the value range.
Max Running Apps	Indicates the maximum number of tasks that can be executed at the same time in the current queue. The default value is **-1**, indicating that the number is not limited within the value range (the meaning is the same if the value is empty). Value **0** indicates that tasks cannot be executed. The value ranges from **-1** to **2147483647**.

Parameter	Description
Max Running Apps per User	Indicates the maximum number of tasks that can be executed by each user in the current queue at the same time. The default value is **-1**, indicating that the number is not limited within the value range (the meaning is the same if the value is empty). Value **0** indicates that tasks cannot be executed. The value ranges from **-1** to **2147483647**.
Max Pending Apps	Indicates the maximum number of tasks that can be suspended at the same time in the current queue. The default value is **-1**, indicating that the number is not limited within the value range (the meaning is the same if the value is empty). Value **0** indicates that tasks cannot be suspended. The value ranges from **-1** to **2147483647**.
Resource Allocation Rule	Indicates the rule for allocating resources to different tasks of a user. The rule can be **FIFO** or **FAIR**. If a user submits multiple tasks in the current queue and the rule is **FIFO**, the tasks are executed one by one in sequential order; If the rule is **FAIR**, resources are evenly allocated to all tasks.
Default Resource Label	Indicates that tasks are executed on a node with a specified resource label.
Active	● **ACTIVE**: indicates that the current queue can receive and execute tasks. ● **INACTIVE**: indicates that the current queue can receive but cannot execute tasks. Tasks submitted to the queue are suspended.
Open	● **OPEN**: indicates that the current queue is opened. ● **CLOSED**: indicates that the current queue is closed. Tasks submitted to the queue are rejected.
Migrate Queue Upon Fault	If cross-AZ HA is enabled for a cluster and an AZ is faulty, set **Migrate Queue Upon Fault** to **TRUE** to migrate running queues of the tenant to other AZs.

Step 5 Click **OK**.

----End

9.7.2.3.5 Configuring the Queue Capacity Policy of a Resource Pool

Scenario

After a resource pool is added, you can configure the capacity policy of available resources for Yarn queues so that jobs in the queues can be properly executed in the resource pool.

This section describes how to configure the queue policy on FusionInsight Manager. Tenant queues equipped with the Superior scheduler can use resources in different resource pools.

Prerequisites

- You have logged in to FusionInsight Manager.
- A resource pool has been added.
- The target queue is not associated with the resource pools of other queues except the default resource pool.

Procedure

Step 1 On FusionInsight Manager, choose **Tenant Resources**.

Step 2 Choose **Dynamic Resource Plan**.

Step 3 Click the **Resource Distribution Policy** tab.

Step 4 Select the name of the target cluster from **Cluster** and select a resource pool from **Resource Pool**.

Step 5 Locate the row that contains the target queue in the **Resource Allocation** area, and click **Modify** in the **Operation** column.

Step 6 On the **Resource Configuration Policy** tab of the **Modify Resource Allocation** window, set the resource configuration policy of the queue in the resource pool.

- **Weight**: indicates the resources that a tenant can obtain. Its initial value is the same as the minimum resource percentage.
- **Minimum Resource**: indicates the minimum resources that a tenant can obtain.
- **Maximum Resource**: indicates the maximum resources that a tenant can obtain.
- **Reserved Resource**: indicates the resources that are reserved for the tenant's queues and cannot be lent to other tenants' queues.

Step 7 Click the **User Policy** tab in the **Modify Resource Allocation** window and set the user policy.

NOTE

defaultUser(built-in) indicates that the policy specified for **defaultUser** is used if a user does not have a policy. The default policy cannot be deleted.

- Click **Add User Policy** to add a user policy.
  - **Username**: indicates the name of a user.
  - **Weight**: indicates the resources that the user can obtain.
  - **Max vCores**: indicates the maximum number of virtual cores that the user can obtain.
  - **Max Memory(MB)**: indicates the maximum memory that the user can obtain.
- Click **Modify** in the **Operation** column to modify an existing user policy.
- Click **Clear** in the **Operation** column to delete an existing user policy.

Step 8 Click **OK**.

----End

9.7.2.3.6 Clearing Queue Configurations

Scenario

You can clear the configurations of a queue on FusionInsight MRS Manager when the queue does not need resources of a resource pool or the resource pool needs to be disassociated from the queue. Clearing queue configurations cancels the resource capacity policy of the queue in the resource pool.

Prerequisites

You have changed the default resource pool of the queue to another one. If a queue is to be disassociated from a resource pool, this resource pool cannot serve as the default resource pool of the queue. For details, see **Configuring a Queue**.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Tenant Resources** > **Dynamic Resource Plan**.

Step 3 Select the name of the target cluster from **Cluster** and select a resource pool from **Resource Pool**.

Step 4 Locate the row that contains the target resource name in the **Resource Allocation** area, and click **Clear** in the **Operation** column.

Step 5 In the displayed dialog box, click **OK** to clear the queue configurations from the current resource pool.

----End

9.7.2.4 Managing Global User Policies

Scenario

If a tenant uses a Superior scheduler, you can configure the global policy for users to use the resource scheduler, including:

- Maximum running apps
- Maximum pending apps
- Default queue

Procedure

- Add a policy.
  a. On FusionInsight Manager, choose **Tenant Resources**.
  b. Choose **Dynamic Resource Plan**.
  c. Click the **Global User Policy** tab.

NOTE

defaults(default setting) indicates that the policy specified for **defaults** is used if a user does not have a global policy. The default policy cannot be deleted.

d. Click **Create Global User Policy**. In the displayed dialog box, set the following parameters:

- **Cluster**: Select the target cluster.
- **Username**: indicates the user for whom resource scheduling is controlled. Enter an existing username in the current cluster.
- **Max Running Apps**: indicates the maximum number of tasks that the user can run in the current cluster.
- **Max Pending Apps**: indicates the maximum number of tasks that the user can suspend in the current cluster.
- **Default Queue**: indicates the queue of the user. Enter the name of an existing queue in the current cluster.

- Modify a policy.

  a. On FusionInsight Manager, choose **Tenant Resources**.

  b. Choose **Dynamic Resource Plan**.

  c. Click the **Global User Policy** tab.

  d. In the row that contains the desired user policy, click **Modify** in the **Operation** column.

  e. In the displayed dialog box, modify parameters and click **OK**.

- Delete a policy.

  a. On FusionInsight Manager, choose **Tenant Resources**.

  b. Choose **Dynamic Resource Plan**.

  c. Click the **Global User Policy** tab.

  d. In the row that contains the desired user policy, click **Delete** in the **Operation** column.

  In the displayed dialog box, click **OK**.

9.7.3 Using the Capacity Scheduler

9.7.3.1 Creating Tenants

9.7.3.1.1 Adding a Tenant

Scenario

You can create tenants on FusionInsight Manager based on the resource consumption and isolation planning and requirements of services.

Prerequisites

- A tenant name has been planned based on service requirements. The name cannot be the same as that of a role, HDFS directory, or Yarn queue that exists in the current cluster.
- Resources to be allocated to the current tenant have been planned to ensure that the sum of resources of direct sub-tenants at each level does not exceed the resources of the current tenant.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Tenant Resources**.

Step 2 Click ⊕. On the page that is displayed, configure tenant attributes according to **Table 9-39**.

Table 9-39 Tenant parameters

Parameter	Description
Cluster	Indicates the cluster for which you want to create a tenant.
Name	● Indicates the name of the current tenant. The value consists of 3 to 50 characters, including digits, letters, and underscores (_). ● Plan a tenant name based on service requirements. The name cannot be the same as that of a role, HDFS directory, or Yarn queue that exists in the current cluster.
Tenant Resource Type	Specifies whether the tenant is a leaf tenant. ● When **Leaf Tenant Resource** is selected, the current tenant is a leaf tenant and no sub-tenant can be added. ● When **Non-leaf Tenant Resource** is selected, the current tenant is not a leaf tenant and sub-tenants can be added to the current tenant.

Parameter	Description
Computing Resource	Specifies the dynamic computing resources for the current tenant. ● When **Yarn** is selected, the system automatically creates a queue in Yarn and the queue is named the same as the tenant name. – A leaf tenant can directly submit jobs to the queue. – A non-leaf tenant cannot directly submit jobs to the queue. However, Yarn adds an extra queue (hidden) named **default** for the non-leaf tenant to record the remaining resource capacity of the tenant. Actual jobs do not run in this queue. ● If **Yarn** is not selected, the system does not automatically create a queue.
Configuration Mode	Indicates the configuration mode of computing resource parameters. ● If you select **Basic**, you only need to set **Default Resource Pool Capacity (%)**. ● If you select **Advanced**, you can manually configure the resource allocation weight and the minimum, maximum, and reserved resources of the tenant.
Default Resource Pool Capacity (%)	Indicates the percentage of computing resources used by the current tenant in the default resource pool. The value ranges from **0** to **100%**.
Weight	Indicates the resource allocation weight. The value ranges from **0** to **100**.
Minimum Resource	Indicates the resources guaranteed for the tenant (preemption supported). The value can be a percentage or an absolute value of the parent tenant's resources. When a tenant has a light workload, the resources of the tenant are automatically allocated to other tenants. When the available tenant resources are less than the value of **Minimum Resource**, the tenant can preempt the resources that have been lent to other tenants.
Maximum Resource	Indicates the maximum resources that can be used by the tenant. The tenant cannot obtain more resources than the value configured. The value can be a percentage or an absolute value of the parent tenant's resources.

Parameter	Description
Reserved Resource	Indicates the resources reserved for the tenant. The reserved resources cannot be used by other tenants even if no job is running in the current tenant resources. The value can be a percentage or an absolute value of the parent tenant's resources.
Storage Resource	Specifies storage resources for the current tenant. ● When **HDFS** is selected, the system automatically allocates storage resources. ● When **HDFS** is not selected, the system does not automatically allocate storage resources.
Quota	Indicates the quota for files and directories.
Space Quota	Indicates the quota for the HDFS storage space used by the current tenant. ● If the unit is set to **MB**, the value ranges from **1** to **8796093022208**. If the unit is set to **GB**, the value ranges from **1** to **8589934592**. ● This parameter indicates the maximum HDFS storage space that can be used by the tenant, but not the actual space used. ● If its value is greater than the size of the HDFS physical disk, the maximum space available is the full space of the HDFS physical disk.
Storage Path	Indicates the HDFS storage directory for the tenant. ● The system automatically creates a folder named after the tenant name in the **/tenant** directory by default. For example, the default HDFS storage directory for tenant **ta1** is **/tenant/ta1**. ● When a tenant is created for the first time, the system creates the **/tenant** directory in the HDFS root directory. The storage path is customizable.
Service	Specifies whether to associate resources of other services. For details, see **Step 4**.
Description	Indicates the description of the current tenant.

NOTE

Roles, computing resources, and storage resources are automatically created when tenants are created.

- The new role has permissions on the computing and storage resources. This role and its permissions are automatically controlled by the system and cannot be manually managed by choosing **System** > **Permission** > **Role**. The role name is in the format of *Tenant name_Cluster ID*. The ID of the first cluster is not displayed by default.
- When using this tenant, create a system user and bind the user to the role of the tenant. For details, see **Adding a User and Binding the User to a Tenant Role**.
- During the tenant creation, the system automatically creates a Yarn queue named after the tenant. If the queue name already exists, the new queue is named **Tenant name**-*N*. *N* indicates a natural number starting from **1**. When a same name exists, the value *N* increases automatically to differentiate the queue from others. For example, **saletenant**, **saletenant-1**, and **saletenant-2**.

Step 3 Check whether the current tenant needs to be associated with resources of other services.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 Click **Associate Service** to configure other service resources used by the current tenant.

- Set **Service** to **HBase** and **Association Type** to **Exclusive** or **Shared**.

  NOTE

  – **Exclusive** indicates that the service resources are used by the tenant exclusively and cannot be associated with other tenants.

  – **Shared** indicates that the service resources can be shared with other tenants.

NOTE

- Only HBase can be associated with a new tenant. However, HDFS, HBase, and Yarn can be associated with existing tenants.
- To associate an existing tenant with service resources, click the target tenant in the tenant list, switch to the **Service Associations** page, and click **Associate Service** to configure resources to be associated with the tenant.
- To disassociate an existing tenant from service resources, click the target tenant in the tenant list, switch to the **Service Associations** page, and click **Delete** in the **Operation** column. In the displayed dialog box, select **I have read the information and understand the impact** and click **OK**.

1. Click **OK**.

Step 5 Click **OK**. Wait until the system displays a message indicating that the tenant is successfully created.

----End

9.7.3.1.2 Adding a Sub-Tenant

Scenario

You can create sub-tenants on FusionInsight Manager and allocate resources of the current tenant to the sub-tenants based on the resource consumption and isolation planning and requirements of services.

Prerequisites

- A parent non-leaf tenant has been added.
- A tenant name has been planned based on service requirements. The name cannot be the same as that of a role, HDFS directory, or Yarn queue that exists in the current cluster.
- Resources to be allocated to the current tenant have been planned to ensure that the sum of resources of direct sub-tenants at each level does not exceed the resources of the current tenant.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Tenant Resources**.

Step 2 In the tenant list on the left, select a parent tenant and click ⊕. On the page for adding a sub-tenant, set attributes for the sub-tenant according to **Table 9-40**.

Table 9-40 Sub-tenant parameters

Parameter	Description
Cluster	Indicates the cluster to which the parent tenant belongs.
Parent Tenant Resource	Indicates the name of the parent tenant.
Name	● Indicates the name of the current tenant. The value consists of 3 to 50 characters, including digits, letters, and underscores (_). ● Plan a sub-tenant name based on service requirements. The name cannot be the same as that of a role, HDFS directory, or Yarn queue that exists in the current cluster.
Tenant Type	Specifies whether the tenant is a leaf tenant. ● When **Leaf Tenant** is selected, the current tenant is a leaf tenant and no sub-tenant can be added. ● When **Non-leaf Tenant** is selected, the current tenant is not a leaf tenant and sub-tenants can be added to the current tenant. However, the tenant depth cannot exceed 5 levels.

Parameter	Description
Computing Resource	Specifies the dynamic computing resources for the current tenant. ● When **Yarn** is selected, the system automatically creates a queue in Yarn and the queue is named the same as the sub-tenant name. – A leaf tenant can directly submit jobs to the queue. – A non-leaf tenant cannot directly submit jobs to the queue. However, Yarn adds an extra queue (hidden) named **default** for the non-leaf tenant to record the remaining resource capacity of the tenant. Actual jobs do not run in this queue. ● If **Yarn** is not selected, the system does not automatically create a queue.
Default Resource Pool Capacity (%)	Indicates the percentage of computing resources used by the current tenant. The base value is the total resources of the parent tenant.
Default Resource Pool Max Capacity (%)	Indicates the maximum percentage of computing resources used by the current tenant. The base value is the total resources of the parent tenant.
Storage Resource	Specifies storage resources for the current tenant. ● When **HDFS** is selected, the system automatically creates a folder named after the sub-tenant in the HDFS parent tenant directory. ● When **HDFS** is not selected, the system does not automatically allocate storage resources.
Quota	Indicates the quota for files and directories.
Space Quota	Indicates the quota for the HDFS storage space used by the current tenant. ● If the unit is set to **MB**, the value ranges from **1** to **8796093022208**. If the unit is set to **GB**, the value ranges from **1** to **8589934592**. ● This parameter indicates the maximum HDFS storage space that can be used by the tenant, but not the actual space used. ● If its value is greater than the size of the HDFS physical disk, the maximum space available is the full space of the HDFS physical disk. ● If this quota is greater than the quota of the parent tenant, the actual storage space does not exceed the quota of the parent tenant.

Parameter	Description
Storage Path	Indicates the HDFS storage directory for the tenant. ● The system automatically creates a folder named after the sub-tenant name in the directory of the parent tenant by default. For example, if the sub-tenant is **ta1s** and the parent directory is **/tenant/ta1**, the storage path for the sub-tenant is then **/tenant/ta1/ta1s**. ● The storage path is customizable in the parent directory.
Description	Indicates the description of the current tenant.

NOTE

Roles, computing resources, and storage resources are automatically created when tenants are created.

- The new role has permissions on the computing and storage resources. This role and its permissions are automatically controlled by the system and cannot be manually managed by choosing **System** > **Permission** > **Role**. The role name is in the format of *Tenant name_Cluster ID*. The ID of the first cluster is not displayed by default.
- When using this tenant, create a system user and bind the user to the role of the tenant. For details, see **Adding a User and Binding the User to a Tenant Role**.
- The sub-tenant can further allocate the resources of its parent tenant. The sum of the resource percentages of direct sub-tenants under a parent tenant at each level cannot exceed 100%. The sum of the computing resource percentages of all level-1 tenants cannot exceed 100%.

Step 3 Check whether the current tenant needs to be associated with resources of other services.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 Click **Associate Service** to configure other service resources used by the current tenant.

1. Set **Services** to **HBase**.
2. Set **Association Type** as follows:
   - **Exclusive** indicates that the service resources are used by the tenant exclusively and cannot be associated with other tenants.
   - **Shared** indicates that the service resources can be shared with other tenants.

NOTE

- Only HBase can be associated with a new tenant. However, HDFS, HBase, and Yarn can be associated with existing tenants.
- To associate an existing tenant with service resources, click the target tenant in the tenant list, switch to the **Service Associations** page, and click **Associate Service** to configure resources to be associated with the tenant.
- To disassociate an existing tenant from service resources, click the target tenant in the tenant list, switch to the **Service Associations** page, and click **Delete** in the **Operation** column. In the displayed dialog box, select **I have read the information and understand the impact** and click **OK**.

3. Click **OK**.

Step 5 Click **OK**. Wait until the system displays a message indicating that the tenant is successfully created.

----End

9.7.3.1.3 Adding a User and Binding the User to a Tenant Role

Scenario

A newly created tenant cannot directly log in to the cluster to access resources. You need to add a user for the tenant on FusionInsight Manager and bind the user to the role of the tenant to assign operation permissions to the user.

Prerequisites

You have clarified service requirements and created a tenant.

Procedure

Step 1 Log in to FusionInsight Manager and choose **System** > **Permission** > **User**.

Step 2 If you want to add a user to the system, click **Create**.

If you want to bind tenant roles to an existing user in the system, locate the row of the user and click **Modify** in the **Operation** column.

Set user attributes according to **Table 9-41**.

Table 9-41 User parameters

Parameter	Description
Username	Specifies the current user name. The value can contain 3 to 32 characters, including digits, letters, underscores (_), hyphens (-), and spaces. ● The username cannot be the same as the OS username of any node in the cluster. Otherwise, the user cannot be used. ● A username that differs only in alphabetic case from an existing username is not allowed. For example, if **User1** has been created, you cannot create **user1**. Enter the correct username when using **User1**.
User Type	The options are **Human-Machine** and **Machine-Machine**. ● **Human-Machine** user: used for FusionInsight Manager O&M and component client operations. If you select this option, set both **Password** and **Confirm Password** accordingly. ● **Machine-Machine** user: used for application development. If you select this option, the password is randomly generated.
Password	This parameter is mandatory if **User Type** is set to **Human-Machine**. The password must contain 8 to 64 characters of at least four types of the following: uppercase letters, lowercase letters, digits, special characters, and spaces. The password cannot be the username or the username spelled backwards.
Confirm Password	Enter the password again.
User Group	In the **User Group** area, click **Add** and select user groups to add the user to the groups. ● If roles have been added to the user groups, the user can be granted the permissions of the roles. ● For example, add the user to the Hive user group to assign Hive permissions to the user.
Primary Group	Select a group as the primary group for the user to create directories and files. The drop-down list contains all groups selected in **User Group**.
Role	Click **Add** to bind a tenant role to the user. **NOTE** If a user wants to use the resources of tenant **tenant1** and to add or delete sub-tenants for **tenant1**, the user must be bound to both the **Manager_tenant** and **tenant1**_*Cluster ID* roles.

Parameter	Description
Description	Indicates the description of the current user.

Step 3 Click **OK**.

----End

9.7.3.2 Managing Tenants

9.7.3.2.1 Managing Tenant Directories

Scenario

You can manage the HDFS storage directories used by specified tenants based on service requirements on FusionInsight Manager, such as adding tenant directories, changing the quotas for directories and files and for storage space, and deleting directories.

Prerequisites

A tenant with HDFS storage resources has been added.

Viewing a Tenant Directory

Step 1 On FusionInsight Manager, choose **Tenant Resources**.

Step 2 In the tenant list on the left, click the target tenant.

Step 3 Click the **Resource** tab.

Step 4 View the **HDFS Storage** table.

- The **File Number Threshold** column provides the quota for files and directories of the tenant directory.
- The **Space Quota** column provides the storage space size of the tenant directory.

----End

Adding a Tenant Directory

Step 1 On FusionInsight Manager, choose **Tenant Resources**.

Step 2 In the tenant list on the left, click the target tenant.

Step 3 Click the **Resource** tab.

Step 4 In the **HDFS Storage** area, click **Create Directory**.

- **Parent Directory**: indicates the storage directory used by the parent tenant of the current tenant.

  NOTE

  This parameter is not displayed if the current tenant is not a sub-tenant.

- Set **Path** to a tenant directory path.

  NOTE

  If the current tenant is not a sub-tenant, the new path is created in the HDFS root directory.

- Set **Quota** to the quota for files and directories.
- **File Number Threshold (%)** is valid only when **Quota** is set. If the ratio of the number of used files to the value of **Quota** exceeds the value of this parameter, an alarm is generated. If this parameter is not specified, no alarm is reported in this scenario.

  NOTE

  The number of used files is collected every hour. Therefore, the alarm indicating that the ratio of used files exceeds the threshold is delayed.

- Set **Space Quota** to the storage space size of the tenant directory.
- If the ratio of used storage space to the value of **Space Quota** exceeds the **Storage Space Threshold (%)** value, an alarm is generated. If this parameter is not specified, no alarm is reported in this scenario.

  NOTE

  The used storage space is collected every hour. Therefore, the alarm indicating that the ratio of used storage space exceeds the threshold is delayed.

Step 5 Click **OK**.

----End

Modifying a Tenant Directory

Step 1 On FusionInsight Manager, choose **Tenant Resources**.

Step 2 In the tenant list on the left, click the target tenant.

Step 3 Click the **Resource** tab.

Step 4 In the **HDFS Storage** table, click **Modify** in the **Operation** column of the specified tenant directory.

- Set **Quota** to the quota for files and directories.
- **File Number Threshold (%)** is valid only when **Quota** is set. If the ratio of the number of used files to the value of **Quota** exceeds the value of this parameter, an alarm is generated. If this parameter is not specified, no alarm is reported in this scenario.
- Set **Space Quota** to the storage space size of the tenant directory.
- If the ratio of used storage space to the value of **Space Quota** exceeds the **Storage Space Threshold (%)** value, an alarm is generated. If this parameter is not specified, no alarm is reported in this scenario.

Step 5 Click **OK**.

----End

Deleting a Tenant Directory

Step 1 On FusionInsight Manager, choose **Tenant Resources**.

Step 2 In the tenant list on the left, click the target tenant.

Step 3 Click the **Resource** tab.

Step 4 In the **HDFS Storage** table, click **Delete** in the **Operation** column of the specified tenant directory.

> **NOTE**
>
> The tenant directory that is created by the system during tenant creation cannot be deleted.

Step 5 Click **OK**.

----End

9.7.3.2.2 Restoring Tenant Data

Scenario

Tenant data is stored on FusionInsight Manager and cluster components. When components are recovered from failures or reinstalled, some configuration data of all tenants may become abnormal. In this case, you need to manually restore the configuration data on FusionInsight Manager.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Tenant Resources**.

Step 2 In the tenant list on the left, click the target tenant.

Step 3 Check the tenant data status.

1. On the **Summary** page, check **Tenant Status**. A green icon indicates that the tenant is available and gray indicates that the tenant is unavailable.
2. Click **Resource** and check the icons on the left of **Yarn** and **HDFS Storage**. A green icon indicates that the resource is available, and gray indicates that the resource is unavailable.
3. Click **Service Associations** and check the **Status** column of the associated services. **Normal** indicates that the component can provide services for the associated tenant. **Not Available** indicates that the component cannot provide services for the tenant.
4. If any of the preceding check items is abnormal, go to **Step 4** to restore tenant data.

Step 4 Click . In the displayed dialog box, enter the password of the current login user and click **OK**.

Step 5 In the **Restore Tenant Resource Data** window, select one or more components to restore data, and click **OK**. The system automatically restores the tenant data.

----End

9.7.3.2.3 Deleting a Tenant

Scenario

You can delete tenants that are no longer used on FusionInsight Manager based on service requirements to release resources occupied by the tenants.

Prerequisites

- A tenant has been added.
- The tenant has no sub-tenants. If the tenant has sub-tenants, delete them; otherwise, the tenant cannot be deleted.
- The role of the tenant is not associated with any user or user group.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Tenant Resources**.

Step 2 In the tenant list on the left, click the target tenant and click 🗑.

NOTE

- If you want to retain the tenant data, select **Reserve the data of this tenant resource**. Otherwise, the storage space of the tenant will be deleted.
- To delete a tenant without retaining the tenant data as a user who does not belong to the supergroup, you should first log in to the HDFS client as a user who belongs to the supergroup and then manually clear the storage space of that tenant to avoid residual data.

Step 3 Click **OK**.

It takes a few minutes to save the configuration. After the tenant is deleted, the role and storage space of the tenant are also deleted.

NOTE

After the tenant is deleted, the queue of the tenant still exists in Yarn. The queue of the tenant is not displayed on the role management page in Yarn.

----End

9.7.3.2.4 Clearing Non-associated Queues of a Tenant

Scenario

If Yarn uses the Capacity scheduler, deleting a tenant only sets the queue capacity of the tenant to **0** and the tenant status to **STOPPED** but does not clear the queues of the tenant in Yarn. Limited by the Yarn mechanism, queues cannot be dynamically deleted. You can run commands to manually delete residual queues.

Impact on the System

- During the script execution, the Controller service is restarted, Yarn configurations are synchronized, and the active and standby ResourceManagers are restarted.
- FusionInsight Manager becomes inaccessible during the restart of the Controller service.

- After the active and standby ResourceManagers are restarted, an alarm is generated indicating that Yarn and components that depend on Yarn are temporarily unavailable.

Prerequisites

Queues of a deleted tenant still exist.

Procedure

Step 1 Check that queues of the deleted tenant still exist.

1. On FusionInsight Manager, choose **Cluster**, click the name of the target cluster, and choose **Services** > **Yarn**. Click the link of the active ResourceManager in **ResourceManager WebUI** to go to the ResourceManager web UI.
2. Click **Scheduler** in the navigation tree on the left. In the right pane, you can view that queues of the tenant still exist in the **STOPPED** state and their **Configured Capacity** is **0**.

Step 2 Log in to the active management node as user **omm**.

Step 3 Switch the directory and execute the **cleanQueuesAndRestartRM.sh** script.

cd ${BIGDATA_HOME}/om-server/om/sbin

./cleanQueuesAndRestartRM.sh -c *Cluster ID*

NOTE

You can choose **Cluster**, click the cluster name, and choose **Cluster Properties** on FusionInsight Manager to view the cluster ID.

During the script execution, you need to enter **yes** and the password.

```
Running the script will restart Controller and restart ResourceManager.
Are you sure you want to continue connecting (yes/no)?yes
Please input admin password:
Begin to backup queues ...
...
```

Step 4 After the script is executed successfully, log in to FusionInsight Manager, choose **Cluster**, click the cluster name, and choose **Services** > **Yarn**. Click the link of the active ResourceManager in **ResourceManager WebUI** to go to the ResourceManager web UI.

Step 5 Click **Scheduler** in the navigation tree on the left. In the right pane, you can view that queues of the tenant have been cleared.

----End

9.7.3.3 Managing Resources

9.7.3.3.1 Adding a Resource Pool

Scenario

In a cluster, you can logically group Yarn NodeManagers into Yarn resource pools. Each NodeManager belongs to only one resource pool. You can create a custom resource pool on FusionInsight Manager and add the hosts that have not been added to any custom resource pools to this resource pool so that specified queues can use the computing resources provided by these hosts.

The system contains a **default** resource pool by default. All NodeManagers that are not added to custom resource pools belong to this resource pool.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Tenant Resources** > **Resource Pool**.

Step 3 Click **Add Resource Pool**.

Step 4 Set resource pool attributes.

- **Cluster**: Select the cluster to which the resource pool is to be added.
- **Name**: Enter the name of the resource pool. The name contains 1 to 50 characters, including digits, letters, and underscores (_), and cannot start with an underscore (_).
- **Resource Label**: Enter the resource label of the resource pool. The value can contain 1 to 50 characters, including digits, letters, underscores (_), and hyphens (-), and must start with a digit or letter.
- **Resource**: In the **Available Hosts** area, select specified hosts and click >> to add the hosts to the **Selected Hosts** area. Only hosts in the cluster can be selected. The host list in the resource pool can be left blank.

  NOTE

  You can filter hosts by host name, number of CPU cores, memory, operating system, or platform type based on service requirements.

Step 5 Click **OK**.

After the resource pool is created, you can view its name, members, and mode in the resource pool list. Hosts that are added to the custom resource pool are no longer members of the **default** resource pool.

----End

9.7.3.3.2 Modifying a Resource Pool

Scenario

When hosts in a resource pool need to be adjusted based on service requirements, you can modify members in the resource pool on FusionInsight Manager.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Tenant Resources** > **Resource Pool**.

Step 3 Locate the row that contains the specified resource pool, and click **Edit** in the **Operation** column.

Step 4 In the **Resource** area, modify hosts.

- Adding hosts: Select desired hosts in **Available Hosts** and click >> to add them to the resource pool.
- Deleting hosts: Select desired hosts in **Selected Hosts** and click << to remove them from the resource pool. The host list in the resource pool can be left blank.

Step 5 Click **OK**.

----End

9.7.3.3.3 Deleting a Resource Pool

Scenario

If a resource pool is no longer used based on service requirements, you can delete it on FusionInsight Manager.

Prerequisites

- Any queue in the cluster does not use the resource pool to be deleted as the default resource pool. Before deleting the resource pool, cancel the default resource pool. For details, see **Configuring a Queue**.
- Resource distribution policies of all queues have been cleared from the resource pool to be deleted. For details, see **Clearing Queue Configurations**.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Tenant Resources** > **Resource Pool**.

Step 3 Locate the row that contains the specified resource pool, and click **Delete** in the **Operation** column.

Step 4 In the displayed dialog box, click **OK**.

----End

9.7.3.3.4 Configuring a Queue

Scenario

You can modify the queue configurations for a specified tenant on FusionInsight Manager.

Prerequisites

A tenant who uses the Capacity scheduler has been added.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Tenant Resources** > **Dynamic Resource Plan**.

The **Resource Distribution Policy** page is displayed by default.

Step 3 Click the **Queue Configurations** tab.

Step 4 Set **Cluster** to the name of the target cluster. In **All tenants resources** area, locate the row that contains the target tenant resource and click **Modify** in the **Operation** column.

NOTE

- You can also access the **Modify Queue Configuration** page as follows: In the tenant list on the **Tenant Resources Management** page, click the target tenant, click the **Resource** tab, and click ✎ next to **Queue Configurations (***Queue name***)**.
- A queue can be bound to only one non-default resource pool. That is, a newly added resource pool can be bound to only one queue to serve as the default resource pool of the queue.

Table 9-42 Queue configuration parameters

Parameter	Description
Tenant Resources Name (Queue)	Indicates the tenant name and queue name.
Maximum Applications	Indicates the maximum number of applications.
Maximum AM Resource Percent	Indicates the maximum percentage of resources that can be used to run the ApplicationMaster in a cluster.
Minimum User Resource Upper-Limit Percent (%)	Indicates the minimum resource guarantee (percentage) of a user. The resources for each user in a queue are limited at any time. If applications of multiple users are running at the same time in a queue, the resource usage of each user fluctuates between the minimum value and the maximum value. The minimum value is determined by the number of running applications, while the maximum value is determined by this parameter. For example, assume that this parameter is set to **25**. If two users submit applications to the queue, each user can use a maximum of 50% resources; if three users submit applications to the queue, each user can use a maximum of 33% resources; if four users submit applications to the queue, each user can use a maximum of 25% resources.

Parameter	Description
User Resource Upper-Limit Factor	Indicates the limit factor of the maximum user resource usage. The maximum user resource usage percentage can be obtained by multiplying the limit factor with the percentage of the tenant's actual resource usage in the cluster.
Status	Indicates the current status of a resource plan. The value can be **Running** or **Stopped**.
Default Resource Pool	Indicates the resource pool used by the queue. The default value is **default**. If you want to change the resource pool, configure the queue capacity first. For details, see **Configuring the Queue Capacity Policy of a Resource Pool**.

Step 5 Click **OK**.

----End

9.7.3.3.5 Configuring the Queue Capacity Policy of a Resource Pool

Scenario

After a resource pool is added, you can configure the capacity policy of available resources for Yarn queues so that jobs in the queues can be properly executed in the resource pool. A queue can have the queue capacity policy of only one resource pool.

You can view queues and configure queue capacity policies in any resource pool. After the queue policies are configured, Yarn queues are associated with resource pools.

Prerequisites

A queue has been added, that is, a tenant associated with computing resources has been created.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Tenant Resources** > **Dynamic Resource Plan**.

The **Resource Distribution Policy** page is displayed by default.

Step 3 Select the name of the target cluster from **Cluster** and select a resource pool from **Resource Pool**.

Step 4 Locate the row that contains the target resource name in the **Resource Allocation** area, and click **Modify** in the **Operation** column.

Step 5 In the **Modify Resource Allocation** window, configure the resource capacity policy of the queue in the resource pool.

- **Capacity (%)**: indicates the percentage of computing resources used by the current tenant.
- **Maximum Capacity (%)**: indicates the maximum percentage of computing resources used by the current tenant.

Step 6 Click **OK**.

NOTE

After the resource capacity values of a queue are deleted and saved, the resource capacity policy of the queue in the resource pool is canceled, indicating that the queue is disassociated from the resource pool. To achieve this, you need to change the default resource pool of the queue to another one. For details, see **Configuring a Queue**.

----End

9.7.3.3.6 Clearing Queue Configurations

Scenario

You can clear the configurations of a queue on FusionInsight MRS Manager when the queue does not need resources of a resource pool or the resource pool needs to be disassociated from the queue. Clearing queue configurations cancels the resource capacity policy of the queue in the resource pool.

Prerequisites

You have changed the default resource pool of the queue to another one. If a queue is to be disassociated from a resource pool, this resource pool cannot serve as the default resource pool of the queue. For details, see **Configuring a Queue**.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Tenant Resources** > **Dynamic Resource Plan**.

Step 3 Select the name of the target cluster from **Cluster** and select a resource pool from **Resource Pool**.

Step 4 Locate the row that contains the target resource name in the **Resource Allocation** area, and click **Clear** in the **Operation** column.

Step 5 In the displayed dialog box, click **OK** to clear the queue configurations from the current resource pool.

----End

9.7.4 Switching the Scheduler

Scenario

The newly installed MRS cluster uses the Superior scheduler by default. If the cluster is upgraded from an earlier version, you can switch the YARN scheduler from the Capacity scheduler to the Superior scheduler with a few clicks.

Prerequisites

- The network connectivity of the cluster is proper and secure, and the YARN service status is normal.
- During scheduler switching, tenants cannot be added, deleted, or modified. In addition, services cannot be started or stopped.

Impact on the System

- Because the ResourceManager is restarted during scheduler switching, submitting jobs to YARN will fail at that time.
- During scheduler switching, tasks in a job being executed on YARN will continue, but new tasks cannot be started.
- After scheduler switching is complete, jobs executed on YARN may fail, causing service interruptions.
- After scheduler switching is complete, parameters of the Superior scheduler are used for tenant management.
- After scheduler switching is complete, tenant queues whose capacity is 0 in the Capacity scheduler cannot be allocated resources in the Superior scheduler. As a result, jobs submitted to these tenant queues fail to be executed. Therefore, you are advised not to set the capacity of a tenant queue to 0 in the Capacity scheduler.
- After scheduler switching is complete, you cannot add or delete resource pools, YARN node labels, or tenants during the observation period. If such an operation is performed, the scheduler cannot be rolled back to the Capacity scheduler.

  NOTE

  - The recommended observation period for scheduler switching is one week. If resource pools, YARN node labels, or tenants are added or deleted during this period, the observation period ends immediately.
- The scheduler rollback may cause the loss of partial or all YARN job information.

Switching from the Capacity Scheduler to the Superior Scheduler

Step 1 Modify YARN service parameters and ensure that the YARN service status is normal.

1. Log in to FusionInsight Manager as an administrator.
2. Log in to FusionInsight Manager and choose **Cluster** > **Services** > **Yarn**. Click **Configurations** then **All Configurations**, search for **yarn.resourcemanager.webapp.pagination.enable**, and check whether the value is **true**.
   - If yes, go to **Step 1.3**.
   - If no, set the parameter to **true** and click **Save** to save the configuration. On the **Dashboard** tab page of YARN, choose **More** > **Restart Service**, verify the identity, and click **OK**. After the service is restarted, go to **Step 1.3**.
3. Choose **Cluster** > *Name of the desired cluster* > **Services**, and check whether the YARN service status is normal.

Step 2 Log in to the active management node as user **omm**.

Step 3 Switch the scheduler.

The following switching modes are available:

0: converts the Capacity scheduler configurations into the Superior scheduler configurations and then switches the Capacity scheduler to the Superior scheduler.

1: converts the Capacity scheduler configurations into the Superior scheduler configurations only.

2: switches the Capacity scheduler to the Superior scheduler only.

- Mode **0** is recommended if the cluster environment is simple and the number of tenants is less than 20.

  Run the following command:

  sh ${BIGDATA_HOME}/om-server/om/sbin/switchScheduler.sh -c *Cluster ID* **-m 0**

  NOTE

  You can choose **Cluster**, click the cluster name, and choose **Cluster Properties** on FusionInsight Manager to view the cluster ID.

  ```
  Start to convert Capacity scheduler to Superior Scheduler, clusterId=1
  Start to convert Capacity scheduler configurations to Superior. Please wait...
  Convert configurations successfully.
  Start to switch the Yarn scheduler to Superior. Please wait...
  Switch the Yarn scheduler to Superior successfully.
  ```

- If the cluster environment or tenant information is complex and you need to retain the queue configurations of the Capacity scheduler on the Superior scheduler, it is recommended that you use mode **1** first to convert the Capacity scheduler configurations, check the converted configurations, and then use mode **2** to switch the Capacity scheduler to the Superior scheduler.

  a. Run the following command to convert the Capacity scheduler configurations into the Superior scheduler configurations:

     sh ${BIGDATA_HOME}/om-server/om/sbin/switchScheduler.sh -c *Cluster ID* **-m 1**

     ```
     Start to convert Capacity scheduler to Superior Scheduler, clusterId=1
     Start to convert Capacity scheduler configurations to Superior. Please wait...
     Convert configurations successfully.
     ```

  b. Run the following command to switch the Capacity scheduler to the Superior scheduler:

     sh ${BIGDATA_HOME}/om-server/om/sbin/switchScheduler.sh -c *Cluster ID* **-m 2**

     ```
     Start to convert Capacity scheduler to Superior Scheduler, clusterId=1
     Start to switch the Yarn scheduler to Superior. Please wait...
     Switch the Yarn scheduler to Superior successfully.
     ```

- If you do not need to retain the queue configurations of the Capacity scheduler, use mode **2**.

  a. Log in to FusionInsight Manager and delete all tenants except the default tenant.

  b. On FusionInsight Manager, delete all resource pools except the default resource pool.

Run the following command to switch the Capacity scheduler to the Superior scheduler:

sh ${BIGDATA_HOME}/om-server/om/sbin/switchScheduler.sh -c *Cluster ID* **-m 2**

```
Start to convert Capacity scheduler to Superior Scheduler, clusterId=1
Start to switch the Yarn scheduler to Superior. Please wait...
Switch the Yarn scheduler to Superior successfully.
```

NOTE

You can query the scheduler switching logs on the active management node.

- ${BIGDATA_LOG_HOME}/controller/aos/switch_scheduler.log
- ${BIGDATA_LOG_HOME}/controller/aos/aos.log

----End

9.8 System

9.8.1 Configuring Permissions

9.8.1.1 Managing Users

9.8.1.1.1 Creating a User

Scenario

FusionInsight Manager supports a maximum of 50,000 users (including built-in users). By default, only user **admin** has the highest operation permissions of FusionInsight Manager. You need to create users on FusionInsight Manager and assign operation permissions to the users based on service requirements.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 On the **User** page, click **Create**.

Step 4 Set **Username**. The username can contain digits, letters, underscores (_), hyphens (-), and spaces. It is case-insensitive and cannot be the same as any existing username in the system or OS.

Step 5 Set **User Type** to **Human-Machine** or **Machine-Machine**.

- **Human-Machine** user: used for FusionInsight Manager O&M and component client operations. If you select this option, you also need to select the password policy and set **Password** and **Confirm Password**.
- **Machine-Machine** user: used for component application development. If you select this option, the password is randomly generated.

Step 6 In the **User Group** area, click **Add** to add one or more user groups to the list.

NOTE

- If the selected user group has been bound to a role or a permission policy has been configured in Ranger, the user can obtain the corresponding permissions.
- After FusionInsight Manager is installed, some user groups generated by default have special permissions. Select desired user groups based on the descriptions on the UI.
- If existing user groups cannot meet your requirements, click **Create User Group** to create a user group. For details, see **Creating a User Group**.

Step 7 Select a group from the **Primary Group** drop-down list to create directories and files.

The drop-down list contains all groups selected in **User Group**.

NOTE

A user can belong to multiple groups (including the primary group and secondary groups). The primary group is set to facilitate maintenance and comply with the permission mechanism of the Hadoop community. The primary group has the same permission control functionality as other groups.

Step 8 In the **Role** area, click **Add** to bind roles to the user.

NOTE

- Adding a role when you create a user can specify the user permissions.
- If the permissions granted to the user from the user group cannot meet service requirements, you can bind other created roles to the user. You can also click **Create Role** to create a role first. For details, see **Creating a Role**.

  It takes 3 minutes to make role permission assignment to the user take effect. If the permissions obtained from the user group are enough, you do not need to add a role.
- After Ranger authentication is enabled for a component, you need to configure Ranger policies to assign permissions to the user except the permissions of default user group or role.
- If a user is not added to a user group or assigned a role, the user cannot view information or perform operations after logging in to FusionInsight Manager.

Step 9 Enter information in **Description**.

Step 10 Click **OK**.

After a human-machine user is created, you need to change the initial password as prompted after logging in to FusionInsight Manager.

----End

9.8.1.1.2 Modifying User Information

Scenario

You can modify user information on FusionInsight Manager, including the user group, primary group, role permission assignment, and user description.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 Locate the row that contains the target user and click **Modify** in the **Operation** column.

Modify the parameters based on service requirements.

NOTE

It takes three minutes at most for the change of the user group or role permissions to take effect.

MRS 3.1.2 or later:

- Users (except **admin**) cannot modify their own password policies.
- Locked users cannot modify their password policies.
- After the password policy bound to a user is modified, the modification takes effect when the user changes the password next time.
- After the password policy bound to a user is modified, if the remaining password validity period is greater than the password validity period in the new password policy, the password validity period is set to the validity period in the new password policy. If the remaining password validity period is less than the password validity period in the new password policy, the password validity period remains unchanged.

Step 4 Click **OK**.

----End

9.8.1.1.3 Exporting User Information

Scenario

You can export information about all created users on FusionInsight Manager.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 Click **Export All** to export all user information at a time.

The exported user information contains the username, creation time, description, user type (**0** indicates a human-machine account, **1** indicates a machine-machine account), primary group, user group list, and roles bound to the user.

Step 4 Set **Save AS** to **TXT** or **CSV**. Click **OK**.

----End

9.8.1.1.4 Locking a User

Scenario

A user may be suspended for a long period of time due to service changes. For security purposes, you can lock such a user.

You can lock a user in using either of the following methods:

- Automatic locking: You can set **Password Retries** in the password policy to automatically lock the user whose login attempts exceed this parameter value. For details, see **Configuring Password Policies**.
- Manual locking: You manually lock a user.

This section describes how to lock a user manually. Machine-machine users cannot be locked.

Impact on the System

A locked user cannot log in to FusionInsight Manager or perform identity authentication in the cluster. A locked user can be used only after being manually unlocked or the lock time expires.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 Locate the row that contains the target user and click **Lock** in the **Operation** column.

Step 4 In the window that is displayed, select **I have read the information and understand the impact**. Click **OK**.

----End

9.8.1.1.5 Unlocking a User

Scenario

You can unlock a user on FusionInsight Manager if the user has been locked because the number of login attempts exceeds the threshold. Only users created on FusionInsight Manager can be unlocked.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 Locate the row that contains the target user and click **Unlock** in the **Operation** column.

Step 4 In the window that is displayed, select **I have read the information and understand the impact**. Click **OK**.

----End

9.8.1.1.6 Deleting a User

Scenario

Based on service requirements, you can delete system users that are no longer used on FusionInsight Manager.

NOTE

- After a user is deleted, the provisioned ticket granting ticket (TGT) is still valid within 24 hours. The user can use the TGT for security authentication and access the system.
- If a new user has the same name as the deleted user, the new user will inherit all owner permissions of the deleted user. You are advised to determine whether to delete the resources owned by the deleted user based on service requirements, for example, files in HDFS.
- The default user **admin** cannot be deleted.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 Locate the row that contains the target user, click **More**, and select **Delete**.

NOTE

To delete users in batches, select the users at a time and click **Delete**.

Step 4 In the displayed dialog box, click **OK**.

----End

9.8.1.1.7 Changing a User Password

Scenario

For security purposes, the password of a human-machine user must be changed periodically.

If users have the permission to use FusionInsight Manager, they can change their passwords on FusionInsight Manager.

If users do not have the permission to use FusionInsight Manager, they can change their passwords on the client.

Prerequisites

- You have obtained the current password policy.
- The user has installed the client on any node in the cluster and obtained the IP address of the node. The password of the client installation user can be obtained from the administrator.

Changing the Password on FusionInsight Manager

Step 1 Log in to FusionInsight Manager.

Step 2 Move the cursor to the username in the upper right corner of the page.

On the user account drop-down menu, choose **Change Password**.

Step 3 On the displayed page, set **Current Password**, **New Password**, and **Confirm Password**, and click **OK**.

The password must meet the following complexity requirements:

- Contains at least 8 characters.
- Contains at least four types of the following: uppercase letters, lowercase letters, numbers, spaces, and special characters (`~!@#$%^&*()-_=+|[{}];',<.>/\?).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be a common easily-cracked password.
- Cannot be the same as the password used in the latest *N* times. *N* indicates the value of **Repetition Rule** configured in **Configuring Password Policies**.

----End

Changing the Password on the Client

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to switch to the client directory, for example, **/opt/client**:

cd /opt/client

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 Change the user password. This operation takes effect for all servers.

kpasswd *System username*

For example, if you want to change the password of system user **test1**, run the **kpasswd test1** command.

The password must meet the following complexity requirements:

- Contains at least 8 characters.
- Contains at least four types of the following: uppercase letters, lowercase letters, numbers, spaces, and special characters (`~!@#$%^&*()-_=+|[{}];',<.>/\?).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be a common easily-cracked password.
- Cannot be the same as the password used in the latest *N* times. *N* indicates the value of **Repetition Rule** configured in **Configuring Password Policies**.

NOTE

If an error occurs during the running of the **kpasswd** command, try the following operations:

- Stop the SSH session and start it again.
- Run the **kdestroy** command and then run the **kpasswd** command again.

----End

9.8.1.1.8 Initializing a Password

Scenario

If a user forgets the password or the public account password needs to be changed periodically, you can initialize the password on FusionInsight Manager. After the password is initialized, the system user needs to change the password upon first login.

NOTE

This operation applies only to human-machine users.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 Locate the row that contains the target user, click **More**, and select **Initialize Password**. In the displayed dialog box, enter the password of the current login user and click **OK**. In the **Initialize Password** dialog box, click **OK**.

Step 4 Set **New Password** and **Confirm Password**, and click **OK**.

The password must meet the following complexity requirements:

- Contains at least 8 characters.
- Contains at least four types of the following: uppercase letters, lowercase letters, numbers, spaces, and special characters (`~!@#$%^&*()-_=+|[{}];',<.>/\?).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be a common easily-cracked password.
- Cannot be the same as the password used in the latest *N* times. *N* indicates the value of **Repetition Rule** configured in **Configuring Password Policies**.

----End

9.8.1.1.9 Exporting an Authentication Credential File

Scenario

If a user uses a security mode cluster to develop applications, the keytab file of the user needs to be obtained for security authentication. You can export keytab files on FusionInsight Manager.

NOTE

After a user password is changed, the exported keytab file becomes invalid, and you need to export a keytab file again.

Prerequisites

Before downloading the keytab file of a Human-Machine user, the password of the user must be changed at least once on the Manager portal or a client;

otherwise, the downloaded keytab file cannot be used For details, see **Changing a User Password**.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 Locate the row that contains the user whose keytab file needs to be exported, choose **More** > **Download Authentication Credential**, specify the save path after the file is automatically generated, and keep the file properly.

The authentication credential includes the **krb5.conf** file of the Kerberos service.

After the authentication credential file is decompressed, you can obtain the following two files:

- The **krb5.conf** file contains the authentication service connection information.
- The **user.keytab** file contains user authentication information.

----End

9.8.1.2 Managing User Groups

Scenario

FusionInsight Manager supports a maximum of 5000 user groups (including built-in user groups). You can create and manage different user groups based on service scenarios on FusionInsight Manager. A user group is bound to a role to obtain operation permissions. After a user is added to a user group, the user can obtain the operation permissions of the user group. A user group can be used to classify users and manage multiple users.

Prerequisites

- You have learned service requirements and created roles required by service scenarios.
- You have logged in to FusionInsight Manager.

Creating a User Group

Step 1 Choose **System** > **Permission** > **User Group**.

Step 2 Above the user group list, click **Create User Group**.

Step 3 Set **Group Name** and **Description**.

The group name contains 1 to 64 characters, including case-insensitive letters, digits, underscores (_), hyphens (-), and spaces. It cannot be the same as an existing user group name in the system.

Step 4 In the **Role** area, click **Add** to select a role and add it.

NOTE

- For components (except HDFS and Yarn) for which Ranger authorization has been enabled, the permissions of non-default roles on Manager do not take effect. You need to configure Ranger policies to assign permissions to user groups.
- If the resource requests of HDFS and Yarn are beyond the Ranger policies, the ACL rules of the components still take effect.

Step 5 In the **User** area, click **Add** to select a user and add it.

Step 6 Click **OK**.

The user group is created.

----End

Viewing User Group Information

By default, all user groups are displayed in the user group list. You can click the arrow on the left of a user group name to view details about the user group, including the user quantity, specific users, and bound roles of the user group.

Modifying Information About a User Group

Locate the row that contains the target user group, and click **Modify** to modify its information.

Exporting Information About a User Group

Click **Export All** to export all user group information at a time in **TXT** or **CSV** format.

The exported user group information contains the user group name, description, user list, and role list.

Deleting a User Group

Locate the row that contains the target user group, and click **Delete**. To delete multiple user groups in batches, select the target user groups and click **Delete** above the user group list. A user group that contains users cannot be deleted. To delete such a user group, delete all its users by modifying the user group first.

9.8.1.3 Managing Roles

Scenario

FusionInsight Manager supports a maximum of 5000 roles (including system built-in roles but excluding roles automatically created by tenants). Based on different service requirements, you need to create and manage different roles on FusionInsight Manager and perform authorization management for FusionInsight Manager and components using roles.

Prerequisites

- You have learned service requirements.
- You have logged in to FusionInsight Manager.

Creating a Role

Step 1 Choose **System** > **Permission** > **Role**.

Step 2 On the displayed page, click **Create Role** and fill in **Role Name** and **Description**.

The role name consists of 3 to 50 characters, including digits, letters, and underscores (_). It cannot be the same as an existing role name in the system.

Step 3 In the **Configure Resource Permission** area, click the cluster whose permissions are to be added and select service permissions for the role.

When setting permissions for a component, enter a resource name in the search text box in the upper right corner and click the search icon to view the search result.

The search result contains only directories, but not subdirectories. Search by keyword supports fuzzy match and is case-insensitive.

NOTE

- For components (except HDFS and Yarn) for which Ranger authorization has been enabled, the permissions of non-default roles on Manager do not take effect. You need to configure Ranger policies to assign permissions to user groups.
- If the resource requests of HDFS and Yarn are beyond the Ranger policies, the ACL rules of the components still take effect.
- A maximum of 1000 permissions can be set for a component at a time.

Step 4 Click **OK**.

----End

Modifying Role Information

Locate the row that contains the target role and click **Modify**.

Exporting Role Information

Click **Export All** to export all role information at a time in **TXT** or **CSV** format.

The exported role information contains the role name, description, and whether the role is the default role.

Deleting a Role

Locate the row that contains the target role and click **Delete**. To delete multiple roles in batches, select the target roles and click **Delete** above the role list. A role bound to a user cannot be deleted. To delete such a role, disassociate the role from the user by modifying the user first.

Task Example (Creating a Manager Role)

Step 1 Choose **System** > **Permission** > **Role**.

Step 2 On the displayed page, click **Create Role** and fill in **Role Name** and **Description**.

Step 3 In the **Configure Resource Permission** area, click **Manager** and set permissions for the role.

Manager permissions:

- Cluster
  - **view** permission: permission to view information on the **Cluster** page and view alarms and events under **O&M** > **Alarm**.
  - **management** permission: permission for management on the **Cluster** and **O&M** pages.
- User
  - **view** permission: permission to view information on pages under **System** > **Permission**.
  - **management** permission: permission for management on pages under **System** > **Permission**.
- Audit

  management permission: permission for management on the **Audit** page.
- Tenant

  management permission: permission for management on the **Tenant** page and permission to view alarms and events under **O&M** > **Alarm**.
- System

  management permission: permission for management on all pages except those under **Permission** on the **System** page and permission to view alarms and events under **O&M** > **Alarm**.

Step 4 Click **OK**.

----End

9.8.1.4 Security Policies

9.8.1.4.1 Configuring Password Policies

Scenario

To keep up with service security requirements, you can set password security rules, user login security rules, and user locking rules on FusionInsight Manager.

NOTICE

- Modify password policies based on service security requirements, because they involve user management security. Otherwise, security risks may be incurred.
- Change the user password after modifying the password policy, and then the new password policy can take effect.

Modifying a Password Policy (Versions Earlier Than MRS 3.1.2)

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **Security Policy** > **Password Policy**.

Step 3 Click **Modify** in the **Operation** column and modify the password policy as prompted.

For details about the parameters, see **Table 9-43**.

Table 9-43 Password policy parameters

Parameter	Description
Minimum Password Length	Indicates the minimum number of characters a password contains. The value ranges from **8** to **32**. The default value is **8**.
Character Types	Indicates how many character types in the following types a password can contain at least: uppercase letters, lowercase letters, digits, spaces, and special characters (~`!?,.:;-_'(){}[]/<>@#$%^&*+|\\=). The value can be **4** or **5**. The default value is **4**, which means that a password can contain uppercase letters, lowercase letters, digits, and special characters. If you set the parameter to **5**, a password can contain all the five character types mentioned above.
Password Retries	Indicates the number of consecutive wrong password attempts allowed before the system locks the user. The value ranges from **3** to **30**. The default value is **5**.
User Lock Duration (Min)	Indicates the time period in which a user is locked when the user lockout conditions are met. The value ranges from **5** to **120**. The default value is **5**.
Password Validity Period (Day)	Indicates the validity period of a password. The value ranges from **0** to **90**. **0** indicates that the password is permanently valid. The default value is **90**.
Repetition Rule	Indicates the number of previous passwords that cannot be reused when you change the password. The value ranges from **1** to **5**. The default value is **1**. This policy applies to only human-machine accounts.

Parameter	Description
Password Expiration Notification (Days)	Indicates the number of days in advance users are notified that their passwords are about to expire. After the value is set, if the difference between the cluster time and the password expiration time is smaller than this value, the user receives password expiration notifications. When logging in to FusionInsight Manager, the user will be notified that the password is about to expire and a message is displayed asking the user to change the password. The value ranges from **0** to *X* (*X* must be set to the half of the password validity period and rounded down). Value **0** indicates that no notification is sent. The default value is **5**.
Interval for Deleting Authentication Failure Records (Min)	Indicates the interval of retaining incorrect password attempts. The value ranges from **0** to **1440**. **0** indicates that incorrect password attempts are permanently retained, and **1440** indicates that incorrect password attempts are retained for one day. The default value is **5**.

Step 4 Click **OK** to save the configurations. Change the user password after modifying the password policy, and then the new password policy can take effect.

----End

Adding a Password Policy (MRS 3.1.2 or Later)

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **Security Policy** > **Password Policy**.

Step 3 Click **Add Password Policy** and modify the password policy as prompted.

For details about the parameters, see **Table 9-44**.

Table 9-44 Password policy parameters

Parameter	Description
Password Policy Name	The value is a string of 3 to 32 characters, including case-insensitive letters, digits, underscores (_), and hyphens (-). It cannot start with a hyphen (-).
Minimum Password Length	Indicates the minimum number of characters a password contains. The value ranges from **8** to **32**.

Parameter	Description
Character Types	Indicates how many character types in the following types a password can contain at least: uppercase letters, lowercase letters, digits, spaces, and special characters (~`!?,.:;-_'(){}[]/<>@#$%^&*+|\=). The value can be **4** or **5**. The default value is **4**, which means that a password can contain uppercase letters, lowercase letters, digits, and special characters. If you set the parameter to **5**, a password can contain all the five character types mentioned above.
Password Retries	Indicates the number of consecutive wrong password attempts allowed before the system locks the user. The value ranges from **3** to **30**.
User Lock Duration (Min)	Indicates the time period in which a user is locked when the user lockout conditions are met. The value ranges from **5** to **120**.
Password Validity Period (Day)	Indicates the validity period of a password. The value ranges from **0** to **90**. **0** indicates that the password is permanently valid.
Repetition Rule	Indicates the number of previous passwords that cannot be reused when you change the password. The value ranges from **1** to **5**. The default value is **1**. This policy applies to only human-machine accounts.
Password Expiration Notification (Days)	Indicates the number of days in advance users are notified that their passwords are about to expire. After the value is set, if the difference between the cluster time and the password expiration time is smaller than this value, the user receives password expiration notifications. When logging in to FusionInsight Manager, the user will be notified that the password is about to expire and a message is displayed asking the user to change the password. The value ranges from **0** to *X* (*X* must be set to the half of the password validity period and rounded down). Value **0** indicates that no notification is sent.
Interval for Deleting Authentication Failure Records (Min)	Indicates the interval of retaining incorrect password attempts. The value ranges from **0** to **1440**. **0** indicates that incorrect password attempts are permanently retained, and **1440** indicates that incorrect password attempts are retained for one day.

Step 4 Click **OK** to save the configurations.

A new user uses the default password policy. After a new password policy is created, you can manually select the password policy when creating a user. You can modify the password policy of an existing user. For details, see **Modifying User Information**.

----End

NOTE

A maximum of 32 password policies can be created.

Modifying a Password Policy (MRS 3.1.2 or Later)

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **Security Policy** > **Password Policy**.

Step 3 Click **Modify** in the row that contains the target password policy. On the **Modify Password Policy** page, modify the password policy as prompted.

For details about the parameters, see **Table 9-44**.

Step 4 Click **OK** to save the configurations.

----End

NOTE

- Users (except **admin**) cannot modify their own password policies.
- After the password policy bound to a user is modified, if the remaining password validity period is greater than the password validity period in the new password policy, the password validity period is set to the validity period in the new password policy. If the remaining password validity period is less than the password validity period in the new password policy, the password validity period remains unchanged.

Deleting a Password Policy (MRS 3.1.2 or Later)

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **Security Policy** > **Password Policy**.

Step 3 Click **Delete** in the row that contains the target password policy. In the dialog box that is displayed, click **OK**.

----End

NOTE

The default password policy and the password policy that has been bound to a user cannot be deleted.

9.8.1.4.2 Configuring the Independent Attribute

Scenario

User **admin** or administrators who are bound to the Manager_administrator role can configure the independent attribute on FusionInsight Manager so that common users (all service users in the cluster) can set or cancel their own independent attributes.

After the independent attribute option is toggled on, service users need to log in to the system and set the independent attribute.

Constraints

- Administrators cannot set or cancel the independent attribute of a user.
- Administrators cannot obtain the authentication credentials of independent users.

Prerequisites

You have obtained the required administrator username and password.

Procedure

Toggling On or Off the Independent Attribute

Step 1 Log in to FusionInsight Manager as user **admin** or a user bound to the Manager_administrator role.

Step 2 Choose **System** > **Permission** > **Security Policy** > **Independent Configurations**.

Step 3 Toggle on or off **Independent Attribute**, enter the password as prompted, and click **OK**.

Step 4 After the identity is authenticated, wait until the OMS configuration is modified and click **Finish**.

NOTE

After the independent attribute is disabled:

- A user who has the attribute can cancel it from the drop-down list of the username in the upper right corner of the page. The user cannot set the independent attribute again once it is cancelled. After the attribute is cancelled, existing independent tables will retain the attribute. However, the user cannot create independent tables again.
- Users without this attribute cannot set or cancel the attribute.

Configuring the Independent Attribute

Step 5 Log in to FusionInsight Manager as a service user.

NOTICE

Administrators cannot initialize the password of the user after the independent attribute is set. If the user password is forgotten, the password cannot be retrieved.

User **admin** cannot set the independent attribute.

Step 6 Move the cursor to the username in the upper right corner of the page.

Step 7 Select **Set Independent** or **Cancel Independent**.

NOTE

- If the independent attribute is toggled on and has been set for the service user, **Cancel Independent** is displayed.
- If the independent attribute is toggled on but has been cancelled for the service user, **Set Independent** is displayed.
- If the independent attribute is toggled off but has been set for the service user, **Cancel Independent** is displayed.
- If the independent attribute is toggled off and has been cancelled for the service user, no option related to the independent attribute is displayed.

Step 8 Enter the password as prompted and click **OK**.

Step 9 After the identity is authenticated, click **OK** in the dialog box.

----End

9.8.2 Configuring Interconnections

9.8.2.1 Configuring SNMP Northbound Parameters

Scenario

If users need to view alarms and monitoring data of a cluster on the O&M platform, you can use Simple Network Management Protocol (SNMP) on FusionInsight Manager to report related data to the network management system (NMS).

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Interconnection** > **SNMP**.

Step 3 Toggle on **SNMP Service**.

The SNMP service is disabled by default. ● indicates that the service is enabled.

Step 4 Set interconnection parameters according to **Table 9-45**.

Table 9-45 Interconnection parameters

Parameter	Description
Version	Specifies the version of SNMP, which can be: ● **V2C**: This is an earlier version with low security. ● **V3**: This is a later version with higher security than SNMP V2C. SNMP V3 is recommended.
Local Port	Specifies the local port. The default value is **20000**. The value ranges from **1025** to **65535**.

Parameter	Description
Read Community Name	Specifies the read-only community name. This parameter is available only when **Version** is set to **V2C**.
Write Community Name	Specifies the write community name. This parameter is available only when **Version** is set to **V2C**.
Security Username	Specifies the SNMP security username. This parameter is available only when **Version** is set to **V3**.
Authentication Protocol	Specifies the authentication protocol. This parameter is available only when **Version** is set to **V3**. SHA is recommended.
Authentication Password	Specifies the authentication password. This parameter is available only when **Version** is set to **V3**.
Confirm Password	Used to confirm the authentication password. This parameter is available only when **Version** is set to **V3**.
Encryption Protocol	Specifies the encryption protocol. This parameter is available only when **Version** is set to **V3**. AES256 is recommended.
Encryption Password	Specifies the encryption password. This parameter is available only when **Version** is set to **V3**.
Confirm Password	Used to confirm the encryption password. This parameter is available only when **Version** is set to **V3**.

NOTE

- The value of **Security Username** cannot contain repeated strings with the unit length as a common factor of 64 (such as 1, 2, 4, and 8), for example, **abab** and **abcdabcd**.
- The **Authentication Password** and **Encryption Password** must contain 8 to 16 characters, including at least three types of the following characters: uppercase letters, lowercase letters, digits, and special characters. The two passwords must be different. The two passwords cannot be the same as the security username or the reverse of the security username.
- For security purposes, periodically change the authentication password and encryption password when the SNMP protocol is used.
- If SNMP v3 is used, a security user will be locked after five consecutive authentication failures within 5 minutes. The user will be automatically unlocked 5 minutes later.

Step 5 Click **Create Trap Target** in the **Trap Target** area. In the displayed dialog box, set the following parameters:

- **Target Symbol**: specifies the trap target ID, which is the ID of the NMS or host that receives traps. The value consists of 1 to 255 characters, including letters or digits.
- **Target IP Address Mode**: specifies the mode of the target IP address. The value can be **IPv4** or **IPv6**.
- **Target IP Address**: specifies the target IP address, which can communicate with the management plane IP address of the management node.
- **Target Port**: specifies the port receiving traps. The port number must be consistent with the peer end and ranges from 0 to 65535.
- **Trap Community Name**: This parameter is available only when **Version** is set to **V2C** and is used to report the community name.

Click **OK**.

The **Create Trap Target** dialog box is closed.

Step 6 Click **OK**.

----End

9.8.2.2 Configuring Syslog Northbound Parameters

Scenario

If users need to view alarms and events of a cluster on the unified alarm reporting platform, you can use the Syslog protocol on FusionInsight Manager to report related data to the alarm platform.

NOTICE

If the Syslog protocol is not encrypted, data may be stolen.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Interconnection** > **Syslog**.

Step 3 Toggle on **Syslog Service**.

The Syslog service is disabled by default. ● indicates that the service is enabled.

Step 4 Set northbound parameters according to **Table 9-46**.

Table 9-46 Syslog interconnection parameters

Parameter Area	Parameter	Description
Syslog Protocol	Server IP Address Mode	Specifies the IP address mode of the interconnected server. The value can be **IPV4** or **IPV6**.

Parameter Area	Parameter	Description
	Server IP Address	Specifies the IP address of the interconnected server.
	Server Port	Specifies the port number for interconnection.
	Protocol	Specifies the protocol type. The options are as follows: ● **TCP** ● **UDP**
	Severity Level	Specifies the severity of the reported message. The options are as follows: ● **Emergency** ● **Alert** ● **Critical** ● **Error** ● **Warning** ● **Notice** ● **Informational** (default value) ● **Debug** NOTE **Severity Level** and **Facility** determine the priority of the sent message. **Priority** = **Facility** × 8 + **Severity Level** For details about the values of **Severity Level** and **Facility**, see **Table 9-47**.
	Facility	Specifies the module where the log is generated. For details about the available values of this parameter, see **Table 9-47**. Default value **local use 0 (local0)** is recommended.
	Identifier	Specifies the product ID. The default value is **FusionInsight Manager**. The identifier can contain a maximum of 256 characters, including letters, digits, underscores (_), periods (.), hyphens (-), spaces, and the following special characters: | $ { }

Parameter Area	Parameter	Description
Report Message	Report Format	Specifies the message format of the alarm report. For details, see the help information on the page. The report format can contain a maximum of 1024 characters, including letters, digits, underscores (_), periods (.), hyphens (-), spaces, and the following special characters: | $ { } **NOTE** For details about each field in the report format, see **Table 9-48**.
	Alarm Type	Specifies the type of the alarm to be reported.
	Alarm Severities	Specifies the level of the alarm to be reported.
Uncleared Alarm Reporting	Periodic Uncleared Alarm Reporting	Specifies whether to report uncleared alarms in a specified period. You can toggle on or off the function. The function is toggled off by default.
	Report Interval (min)	Specifies the interval for periodically reporting uncleared alarms. This parameter is valid only when **Periodic Uncleared Alarm Reporting** is enabled. The default value is **15**, in minutes. The value ranges from **5** to **1440** (one day).
Heartbeat Settings	Heartbeat Reporting	Specifies whether to periodically report Syslog heartbeat messages. You can toggle on or off the function. The function is toggled off by default.
	Heartbeat Interval (minutes)	Specifies the interval for periodically reporting heartbeat messages. This parameter is valid only when **Heartbeat Reporting** is enabled. The default value is **15**, in minutes. The value ranges from **1** to **60**.
	Heartbeat Packet	Specifies the heartbeat message to be reported. This parameter is valid when **Heartbeat Reporting** is toggled on and cannot be left blank. The value can contain a maximum of 256 characters, including digits, letters, underscores (_), vertical bars (|), colons (:), spaces, commas (,), and periods (.).

NOTE

After the periodic heartbeat packet function is enabled, packets may be interrupted during automatic recovery of some cluster error tolerance (for example, active/standby OMS switchover). In this case, wait for automatic recovery.

Step 5 Click **OK**.

----End

Related Information

Table 9-47 Numeric codes of **Severity Level** and **Facility**

Severity Level	Facility	Numeric Code
Emergency	kernel messages	0
Alert	user-level messages	1
Critical	mail system	2
Error	system daemons	3
Warning	security/authorization messages (note 1)	4
Notice	messages generated internally by syslog	5
Informational	line printer subsystem	6
Debug	network news subsystem	7
-	UUCP subsystem	8
-	clock daemon (note 2)	9
-	security/authorization messages (note 1)	10
-	FTP daemon	11
-	NTP subsystem	12
-	log audit (note 1)	13
-	log alert (note 1)	14
-	clock daemon (note 2)	15
-	local use 0~7 (local0 ~ local7)	16 to 23

Table 9-48 Report format information fields

Information Field	Description
dn	Cluster name
id	Alarm ID
name	Alam name

Information Field	Description
serialNo	Alarm serial number **NOTE** The serial numbers of the fault alarms and the corresponding clear alarms are the same.
category	Alarm type. The options are as follows: ● **0**: fault alarm ● **1**: clear alarm ● **2**: event
occurTime	Time when the alarm was generated
clearTime	Time when this alarm was cleared
isAutoClear	Whether an alarm is automatically cleared. The options are as follows: ● **1**: yes ● **0**: no
locationInfo	Location where the alarm was generated
clearType	Alarm clearance type. The options are as follows: ● **-1**: not cleared ● **0**: automatically cleared ● **2**: manually cleared
level	Severity. The options are as follows: ● **1**: critical alarm ● **2**: major alarm ● **3**: minor alarm ● **4**: warning alarm
cause	Alarm cause
additionalInfo	Additional information
object	Alarm object

9.8.2.3 Configuring Monitoring Metric Dumping

Scenario

The monitoring data reporting function writes the monitoring data collected in the system into a text file and uploads the file to a specified server in FTP or SFTP mode.

Before using this function, you need to perform related configurations on FusionInsight Manager.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Interconnection** > **Upload Performance Data**.

Step 3 Toggle on **Upload Performance Data**.

The performance data upload service is disabled by default. ● indicates that the service is enabled.

Step 4 Set the upload parameters according to **Table 9-49**.

Table 9-49 Upload parameters

Parameter	Description
FTP IP Address Mode	Specifies the server IP address mode. This parameter is mandatory. The value can be **IPV4** or **IPV6**.
FTP IP Address	Specifies the IP address of the FTP server for storing monitoring files after the monitoring metric data is interconnected. This parameter is mandatory.
FTP Port	Specifies the port for connecting to the FTP server. This parameter is mandatory.
FTP Username	Specifies the username for logging in to the FTP server. This parameter is mandatory.
FTP Password	Specifies the password for logging in to the FTP server. This parameter is mandatory.
Save Path	Specifies the path for storing monitoring files on the FTP server. This parameter is mandatory.
Dump Interval (second)	Specifies the interval at which monitoring files are periodically stored on the FTP server, in seconds. This parameter is mandatory.
Dump Mode	Specifies the protocol used for sending monitoring files. This parameter is mandatory. The value can be **SFTP** or **FTP**. You are advised to use the SFTP mode based on SSH v2. Otherwise, security risks may be incurred.
SFTP Service Public Key	Specifies the public key of the FTP server. This parameter is optional. It is valid only when **Dump Mode** is set to **SFTP**.

Step 5 Click **OK**.

NOTE

If the dump mode is SFTP and the public key of the SFTP service is empty, the system displays a security risk warning. You need to evaluate the security risk and then save the configuration.

----End

Data Format

After the configuration is complete, the monitoring data reporting function periodically writes monitoring data in the cluster to text files and reports the files to the corresponding FTP/SFTP service based on the configured reporting period.

- Principles for generating monitoring files
  - The monitoring metrics are written to files generated every 30, 60, and 300 seconds based on the metric collection period.

    30s: real-time metrics that are collected every 30s by default

    60s: real-time metrics that are collected every 60s by default

    300s: all metrics that are not collected every 30s or 60s
  - File name format: *metric_{Interval}_{File creation time YYYYMMDDHHMMSS}.log*

    Example: **metric_60_20160908085915.log**

    metric_300_20160908085613.log
- Monitoring file content
  - Format of monitoring files:

    "Cluster ID|Cluster name|Displayed name|Service name|Metric ID|Collection time|Collection host@m@Sub-metric|Unit|Metric value", where fields are separated using vertical bars (|). For example:

    ```
    1|xx1|Host|Host|10000413|2019/06/18 10:05:00|189-66-254-146|KB/s|309.910
    1|xx1|Host|Host|10000413|2019/06/18 10:05:00|189-66-254-152|KB/s|72.870
    2|xx2|Host|Host|10000413|2019/06/18 10:05:00|189-66-254-163|KB/s|100.650
    ```

    Note: The actual files are not in that format.
  - Interval for uploading monitoring files:

    The interval for uploading monitoring files can be set using the **Dump Interval (second)** parameter on the page. Currently, the interval can range from **30** to **300**. After the configuration is complete, the system periodically uploads files to the corresponding FTP/SFTP server at the specified interval.
- Monitoring metric description file
  - Metric set file

    The metric set file **all-shown-metric-zh_CN** contains detailed information about all metrics. After obtaining the metric IDs from the files reported by the third-party system, you can query details about the metrics from the metric set file.

    Location of the metric set file:

    Active and standby OMS nodes: {*FusionInsight installation path*} **/om-server/om/etc/om/all-shown-metric-zh_CN**

    Content of the metric set file:

    ```
    Real-Time Metric ID,5-Minute Metric ID,Metric Name,Metric Collection Period (s),Collected by
    Default,Service Belonged To,Role Belonged To
    00101,10000101,JobHistoryServer non-heap memory usage,30,false,Mapreduce,JobHistoryServer
    00102,10000102,JobHistoryServer non-heap memory allocation volume,
    30,false,Mapreduce,JobHistoryServer
    00103,10000103,JobHistoryServer heap memory usage,30,false,Mapreduce,JobHistoryServer
    00104,10000104,JobHistoryServer heap memory allocation volume,
    30,false,Mapreduce,JobHistoryServer
    00105,10000105,Number of blocked threads,30,false,Mapreduce,JobHistoryServer
    ```

```
00106,10000106,Number of running threads,30,false,Mapreduce,JobHistoryServer
00107,10000107,GC time,30,false,Mapreduce,JobHistoryServer
00110,10000110,JobHistoryServer CPU usage,30,false,Mapreduce,JobHistoryServer
...
```

- Field description of critical metrics

  Real-Time Metric ID: indicates the ID of the metric whose collection period is 30s or 60s.

  5-Minute Metric ID: indicates the ID of a 5-minute (300s) metric.

  Metric Collection Period (s): indicates the collection period of real-time metrics. The value can be **30** or **60**.

  Service Belonged To: indicates the name of the service to which a metric belongs, for example, HDFS and HBase.

  Role Belonged To: indicates the name of the role to which a metric belongs, for example, JobServer and RegionServer.

- Description

  For metrics whose collection period is 30s/60s, you can find the corresponding metric description by referring to the first column, that is, **Real-Time Metric ID**.

  For metrics whose collection period is 300s, you can find the corresponding metric description by referring to the second column, that is, **5-Minute Metric ID**.

9.8.3 Importing a Certificate

Scenario

CA certificates are used to encrypt data during communication between FusionInsight Manager modules and between cluster component clients and servers to ensure security. CA certificates can be quickly imported to FusionInsight Manager for product security. Import CA certificates in following scenarios:

- When the cluster is installed for the first time, you need to replace the enterprise certificate.
- If the enterprise certificate has expired or security hardening is required, you need to replace it with a new certificate.

Impact on the System

- During certificate replacement, the cluster needs to be restarted. In this case, the system becomes inaccessible and cannot provide services.
- After the certificate is replaced, the certificates used by all components and FusionInsight Manager modules are automatically updated.
- After the certificate is replaced, you need to reinstall the certificate in the local environment where the certificate is not trusted.

Prerequisites

- You have generated the certificate file and key file or obtained them from the enterprise certificate administrator.
- You have obtained the files to be imported to the cluster, including the CA certificate file (***.crt**), key file (***.key**), and file that saves the key file password

(**password.property**). The certificate name and key name can contain uppercase letters, lowercase letters, and digits. After the preceding files are generated, compress them into a TAR package.

- You have obtained a password for accessing the key file, for example, **Userpwd@123**.

  To avoid potential security risks, the password must meet the following complexity requirements:

  - Contains at least 8 characters.
  - Contains at least four types of the following: uppercase letters, lowercase letters, numbers, and special characters (~`!?,.;-_'(){}[]/<>@#$%^&*+|\=).

- When applying for certificates from the certificate administrator, you have provided the password for accessing the key file and applied for the certificate files in CRT, CER, CERT, and PEM formats and the key files in KEY and PEM formats. The requested certificates must have the issuing function.

Procedure

Step 1 Log in to FusionInsight Manager and choose **System** > **Certificate**.

Step 2 Click ... on the right of **Upload Certificate**. In the file selection window, browse to select the obtained TAR package of the certificate files.

Step 3 Click **Upload**.

Manager uploads the compressed package and automatically imports the package.

Step 4 After the certificate is imported, the system displays a message asking you to synchronize the cluster configuration and restart the web service for the new certificate to take effect. Click **OK**.

Step 5 In the displayed dialog box, enter the password of the current login user and click **OK**. The cluster configuration is automatically synchronized and the web service is restarted.

Step 6 After the cluster is restarted, enter the URL for accessing FusionInsight Manager in the address box of the browser and check whether the FusionInsight Manager web page can be successfully displayed.

Step 7 Log in to FusionInsight Manager.

Step 8 Choose **Cluster**, click the name of the target cluster, choose **Dashboard**, click **More**, and select **Restart**.

Step 9 In the displayed dialog box, enter the password of the current login user and click **OK**.

----End

9.8.4 OMS Management

9.8.4.1 Overview of the OMS Page

Overview

Log in to FusionInsight Manager and choose **System** > **OMS**. You can perform maintenance operations on the OMS page, including viewing basic information, viewing the service status of OMS service modules, and manually triggering health checks.

NOTE

OMS is the management node of the O&M system. Generally, there are two OMS nodes that work in active/standby mode.

Basic Information

OMS-associated information is displayed on FusionInsight Manager, as listed in **Table 9-50**.

Table 9-50 OMS information

Item	Description
Version	Indicates the OMS version, which is consistent with the FusionInsight Manager version.
IP Mode	Indicates the IP address mode of the current cluster network.
HA Mode	Indicates the OMS working mode, which is specified by the configuration file during FusionInsight Manager installation.
Current Active	Indicates the host name of the active OMS node, that is, the host name of the active management node. Click a host name to go to the host details page.
Current Standby	Indicates the host name of the standby OMS node, that is, the host name of the standby management node. Click a host name to go to the host details page.
Duration	Indicates the duration for starting the OMS process.

OMS Service Status

FusionInsight Manager displays the running status of all OMS service modules. If the status of each service module is displayed as 🟢, the OMS is running properly.

Health Check

You can click **Health Check** on the OMS page to check the OMS status. If some check items are faulty, you can view the check description for troubleshooting.

Entering or Exiting Maintenance Mode

Configure OMS to enter or exit the maintenance mode.

System Parameters

Connect to the DMPS cluster in large-scale cluster scenarios.

9.8.4.2 Modifying OMS Service Configuration Parameters

Scenario

Based on the security requirements of the user environment, you can modify the Kerberos and LDAP configurations in the OMS on FusionInsight Manager.

Impact on the System

After the OMS service configuration parameters are modified, the corresponding OMS module needs to be restarted. In this case, FusionInsight Manager cannot be used.

Procedure

Modifying the okerberos configuration

Step 1 Log in to FusionInsight Manager and choose **System** > **OMS**.

Step 2 Locate the row that contains okerberos and click **Modify Configuration**.

Step 3 Modify the parameters according to **Table 9-51**.

Table 9-51 okerberos parameters

Parameter	Description
KDC Timeout (ms)	Timeout duration for an application to connect to Kerberos, in milliseconds. The value must be an integer.
Max Retries	Maximum number of retries for an application to connect to Kerberos, in seconds. The value must be an integer.
LDAP Timeout (ms)	Timeout duration for Kerberos to connect to LDAP, in milliseconds.
LDAP Search Timeout (ms)	Timeout duration for Kerberos to query user information in LDAP, in milliseconds.
Kadmin Listening Port	Port number of the Kadmin service.

Parameter	Description
KDC Listening Port	Port number of the kinit service.
Kpasswd Listening Port	Port number of the Kpasswd service.
Reset LDAP Account Password	Machine-machine users (**cn=krbadmin,ou=Users,dc=hadoop,dc=com** and **cn=krbkdc,ou=Users,dc=hadoop,dc=com**) used by Kerberos to access LDAP. If this parameter is selected, the passwords will be replaced by random passwords. **NOTE** This parameter is available only in MRS 3.1.2 or later.

Step 4 Click **OK**.

In the displayed dialog box, enter the password of the current login user and click **OK**. In the displayed confirmation dialog box, click **OK**.

Modifying the oldap configuration

Step 5 Locate the row that contains the oldap and click **Modify Configuration**.

Step 6 Modify the parameters according to **Table 9-52**.

Table 9-52 OLDAP parameters

Parameter	Description
LDAP Listening Port	Port number of the LDAP service.
Reset LDAP Account Password	Machine-machine users (**cn=root,dc=hadoop,dc=com** and **cn=pg_search_dn,ou=Users,dc=hadoop,dc=com**) used by LDAP for data management, synchronization, and status check. If this parameter is selected, the passwords will be replaced by random passwords. **NOTE** This parameter is available only in MRS 3.1.2 or later.

Step 7 Click **OK**.

In the displayed dialog box, enter the password of the current login user and click **OK**. In the displayed confirmation dialog box, click **OK**.

NOTE

To reset the password of the LDAP account, you need to restart ACS. The procedure is as follows:

1. Log in to the active management node as user **omm** using PuTTY, and run the following command to update the domain configuration:

   sh ${BIGDATA_HOME}/om-server/om/sbin/restart-RealmConfig.sh

   The command is run successfully if the following information is displayed:

   ```
   Modify realm successfully. Use the new password to log into FusionInsight again.
   ```

2. Run the **sh $CONTROLLER_HOME/sbin/acs_cmd.sh stop** command to stop ACS.
3. Run the **sh $CONTROLLER_HOME/sbin/acs_cmd.sh start** command to start ACS.

Restarting the cluster

Step 8 Log in to FusionInsight Manager and restart the cluster by referring to **Performing a Rolling Restart of a Cluster**.

----End

9.8.5 Component Management

9.8.5.1 Viewing Component Packages

Scenario

A complete MRS cluster consists of multiple component packages. Before installing some services on FusionInsight Manager, check whether the component packages of those services have been installed.

Procedure

Step 1 Log in to FusionInsight Manager and choose **System** > **Component**.

Step 2 On the **Installed Component** page, view all components.

NOTE

In the **Platform Type** column, you can view the registered OS and platform type of the component.

Step 3 Click ⌄ on the left of a component name to view the services and version numbers contained in the component.

----End

9.9 Cluster Management

9.9.1 Configuring Client

9.9.1.1 Installing a Client

Scenario

This section describes how to install the clients of all services, except Flume, in the MRS cluster. MRS provides shell scripts for different services so that maintenance personnel can log in to related maintenance clients and implement maintenance operations.

NOTE

- Reinstall the client after server configuration is modified on FusionInsight Manager or after the system is upgraded. Otherwise, the versions of the client and server will be inconsistent.

Prerequisites

- An installation directory will be automatically created if it does not exist. If the directory exists, it must be empty. The directory cannot contain any space.
- If a server outside the cluster is used as the client node, the node can communicate with the cluster service plane. Otherwise, client installation will fail.
- The client must have the NTP service enabled and synchronized time with the NTP server. Otherwise, client installation will fail.
- If clients of all components are downloaded, HDFS and MapReduce are installed in the same directory (*Client directory***/HDFS/**).
- You can install and use the client as any user whose username and password have been obtained from the system administrator. This section uses **user_client** as an example. Ensure that user **user_client** is the owner of the server file directory (for example, **/opt/Bigdata/hadoopclient**) and client installation directory (for example, **/opt/client**). The permission for the two directories is **755**.
- You have obtained the component service username (a default user or new user) and password from the system administrator.
- When you install the client as a user other than **omm** or **root**, and the **/var/tmp/patch** directory already exists, you have changed the permission for the directory to **777** and changed the permission for the logs in the directory to **666**.

Procedure

Step 1 Obtain the required software packages.

Log in to FusionInsight Manager. Click the wanted cluster from the **Cluster** drop-down list.

Click **More** and select **Download Client**. The **Download Cluster Client** page is displayed.

NOTE

If only one component client is to be installed, choose **Cluster**, click the name of the target cluster, choose **Services**, click a service name, click **More**, and select **Download Client**. The **Download Client** page is displayed.

Step 2 Set **Select Client Type** to **Complete Client**.

Configuration Files Only is to download client configuration files in the following scenario: After a complete client is downloaded and installed and the system administrator modifies server configurations on Manager, developers need to update the configuration files during application development.

The platform type can be set to **x86_64** or **aarch64**.

- **x86_64**: indicates the client software package that can be deployed on the x86 servers.
- **aarch64**: indicates the client software package that can be deployed on the TaiShan servers.

NOTE

The cluster supports two types of clients: **x86_64** and **aarch64**. The client type must match the architecture of the node for installing the client. Otherwise, client installation will fail.

Step 3 Determine whether to generate a client file on the cluster node.

- If yes, select **Save to Path**, and click **OK** to generate the client file. By default, the client file is generated in **/tmp/FusionInsight-Client** on the active management node. You can also store the client file in other directories, and user **omm** has the read, write, and execute permissions on the directories. Copy the software package to the file directory, for example, **/opt/Bigdata/hadoopclient**, on the server where the client is to be installed as user **omm** or **root**. Then, go to **Step 5**.

  NOTE

  If you cannot obtain the permissions of user **root**, use user **omm**.

- If no, click **OK** and specify a local save path to download the complete client. Wait until the download is complete and go to **Step 4**.

Step 4 Upload the software package.

Use WinSCP to upload the obtained software package as the user (such as **user_client**) who prepares for the installation, to the directory (such as **/opt/Bigdata/hadoopclient**) of the server where the client is to be installed.

The name of the client software package is in the follow format: **FusionInsight_Cluster_***<Cluster ID>***_Services_Client.tar**.

The following steps and sections use **FusionInsight_Cluster_1_Services_Client.tar** as an example.

NOTE

The host where the client is to be installed can be a node inside or outside the cluster. If the node is a server outside the cluster, it must be able to communicate with the cluster, and the NTP service must be enabled to ensure that the time is the same as that on the server.

For example, you can configure the same NTP clock source for external servers as that of the cluster. After the configuration, you can run the **ntpq -np** command to check whether the time is synchronized.

- If there is an asterisk (*) before the IP address of the NTP clock source in the command output, the synchronization is normal. For example:

```
remote refid st t when poll reach delay offset jitter
==============================================================================
=
*10.10.10.162 .LOCL. 1 u 1 16 377 0.270 -1.562 0.014
```

- If there is no asterisk (*) before the IP address of the NTP clock source and the value of **refid** is **.INIT.**, or if the command output is abnormal, the synchronization is abnormal. Contact technical support.

```
remote refid st t when poll reach delay offset jitter
==============================================================================
=
10.10.10.162 .INIT. 1 u 1 16 377 0.270 -1.562 0.014
```

You can also configure the same chrony clock source for external servers as that for the cluster. After the configuration, run the **chronyc sources** command to check whether the time is synchronized.

- In the command output, if there is an asterisk (*) before the IP address of the chrony service on the active OMS node, the synchronization is normal. For example:

```
MS Name/IP address         Stratum Poll Reach LastRx Last sample
==============================================================================
=
^* 10.10.10.162             10  10   377   626    +16us[  +15us] +/-  308us
```

- In the command output, if there is no asterisk (*) before the IP address of the NTP service on the active OMS node, and the value of **Reach** is **0**, the synchronization is abnormal.

```
MS Name/IP address         Stratum Poll Reach LastRx Last sample
==============================================================================
=
^? 10.1.1.1                  0  10     0     -     +0ns[   +0ns] +/-    0ns
```

Step 5 Log in as user **user_client** to the server where the client is to be installed.

Step 6 Decompress the software package.

Go to the directory where the installation package is stored, for example, **/opt/Bigdata/hadoopclient**. Run the following command to decompress the installation package to a local directory:

tar -xvf FusionInsight_Cluster_1_Services_Client.tar

Step 7 Verify the software package.

Run the following command to verify the decompressed file and check whether the command output is consistent with the information in the **sha256** file:

sha256sum -c FusionInsight_Cluster_1_Services_ClientConfig.tar.sha256

```
FusionInsight_Cluster_1_Services_ClientConfig.tar: OK
```

Step 8 Decompress the obtained installation file.

tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar

Step 9 Configure network connections for the client.

1. Ensure that the host where the client is installed can communicate with the hosts listed in the **hosts** file in the decompression directory (for example, **/opt/Bigdata/hadoopclient/FusionInsight_Cluster_***<Cluster ID>***_Services_ClientConfig/hosts**).
2. If the host where the client is installed is not a host in the cluster, you need to set the mapping between the host name and the service plane IP address for each cluster node in **/etc/hosts**, as user **root**. Each host name uniquely maps an IP address. You can perform the following steps to import the domain name mapping of the cluster to the **hosts** file:

   a. Switch to user **root** or a user who has the permission to modify the **hosts** file.

      su - root

   b. Go to the directory where the client package is decompressed.

      cd /opt/Bigdata/hadoopclient/FusionInsight_Cluster_1_Services_ClientConfig

   c. Run the **cat realm.ini >> /etc/hosts** command to import the domain name mapping to the **hosts** file.

NOTE

- If the host where the client is installed is not a node in the cluster, configure network connections for the client to prevent errors when you run commands on the client.
- If Spark tasks are executed in yarn-client mode, add the **spark.driver.host** parameter to the file *Client installation directory***/Spark/spark/conf/spark-defaults.conf** and set the parameter to the client IP address.
- If the yarn-client mode is used, you need to configure the mapping between the IP address and host name of the client in the **hosts** file on the active and standby Yarn nodes (ResourceManager nodes in the cluster) to make sure that the Spark web UI is properly displayed.

Step 10 Go to the directory where the installation package is stored, and run the following command to install the client to a specified directory (an absolute path), for example, **/opt/client**:

cd /opt/Bigdata/hadoopclient/FusionInsight_Cluster_1_Services_ClientConfig

Run the **./install.sh /opt/client** command to install the client. The client is successfully installed if information similar to the following is displayed:

```
The component client is installed successfully
```

NOTE

- If the **/opt/hadoopclient** directory has been used by existing service clients, you need to use another directory in this step when installing other service clients.
- You must delete the client installation directory when uninstalling a client.
- To ensure that an installed client can only be used by the installation user (for example, **user_client**), add parameter **-o** during the installation. That is, run the **./install.sh /opt/hadoopclient -o** command to install the client.
- If the NTP server is to be installed in **chrony** mode, ensure that the parameter **chrony** is added during the installation, that is, run the **./install.sh /opt/client -o chrony** command to install the client.
- If an HBase client is installed, it is recommended that the client installation directory contain only uppercase and lowercase letters, digits, and special characters (_-?.@+=) due to the limitation of the Ruby syntax used by HBase.
- If the client node is a server outside the cluster and cannot communicate with the service plane IP address of the active OMS node or cannot access port 20029 of the active OMS node, the client can be successfully installed but cannot be registered with the cluster or displayed on the UI.

Step 11 Log in to the client to check whether the client is successfully installed.

1. Run the **cd /opt/client** command to go to the client installation directory.
2. Run the **source bigdata_env** command to configure environment variables for the client.
3. For a cluster in security mode, run the following command to set **kinit** authentication and enter the password for logging in to the client. For a cluster in normal mode, user authentication is not required.

   kinit admin

   ```
   Password for xxx@HADOOP.COM: #Enter the login password of user admin (same as the user
   password for logging in to the cluster).
   ```

4. Run the **klist** command to query and confirm authentication details.

   ```
   Ticket cache: FILE:/tmp/krb5cc_0
   Default principal: xxx@HADOOP.COM

   Valid starting       Expires              Service principal
   04/09/2021 18:22:35  04/10/2021 18:22:29  krbtgt/HADOOP.COM@HADOOP.COM
   ```

   NOTE

   - When kinit authentication is used, the ticket is stored in the **/tmp/krb5cc_***uid* directory by default.

     uid indicates the ID of the user who logs in to the OS. For example, if the UID of user **root** is 0, the ticket generated for kinit authentication after user **root** logs in to the system is stored in the **/tmp/krb5cc_0** directory.

     If the current user does not have the read/write permission for the **/tmp** directory, the ticket cache path is changed to **Client installation directory/tmp/krb5cc_uid**. For example, if the client installation directory is **/opt/hadoopclient**, the kinit authentication ticket is stored in **/opt/hadoopclient/tmp/krb5cc_uid**.
   - If the same user is used to log in to the OS for kinit authentication, there is a risk that tickets are overwritten. You can set the **-c** *cache_name* parameter to specify the ticket cache path or set the **KRB5CCNAME** environment variable to avoid this problem.

Step 12 After the cluster is reinstalled, the previously installed client is no longer available. Perform the following operations to deploy the client again:

1. Log in to the node where the client is deployed as user **root**.

2. Run the following command to view the directory where the client is located: (In the following example, **/opt/hadoopclient** is the directory where the client is located.)

   ll /opt

   ```
   drwxr-x---. 6 root root       4096 Dec 11 19:00 hadoopclient
   drwxr-xr-x. 3 root root       4096 Dec  9 02:04 godi
   drwx------. 2 root root      16384 Nov  6 01:03 lost+found
   drwxr-xr-x. 2 root root       4096 Nov  7 09:49 rh
   ```

3. Run the following command to delete the files in the folder (for example, **/opt/client**) where all client programs are located:

   mv /opt/client */tmp/clientbackup*

4. Reinstall the client.

----End

9.9.1.2 Using a Client

Scenario

After the client is installed, you can use the shell command on the client in O&M or service scenarios, or use the sample project on the client during application development.

This section describes how to use the client in O&M scenario or service scenarios.

Prerequisites

- You have installed the client.

  For example, the installation directory is **/opt/client**.

- Service users of each component have been created by the system administrator based on service requirements.

  A machine-machine user needs to download the **keytab** file and a human-machine user needs to change the password upon the first login.

Procedure

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to switch to the client installation directory:

cd /opt/client

Step 3 Run the following command to set environment variables:

source bigdata_env

Step 4 If the cluster is in security mode, authenticate the user. For a normal cluster, user authentication is not required.

kinit *Component service user*

Step 5 Run the **shell** command as required.

----End

9.9.1.3 Updating the Configuration of an Installed Client

Scenario

The cluster provides a client for you to connect to a server, view task results, or manage data. If you modify service configuration parameters on FusionInsight Manager and restart the service, you need to download and install the installed client again or use the configuration file to update the client.

Prerequisites

You have installed a client.

Procedure

Method 1:

Step 1 Log in to FusionInsight Manager. Click the wanted cluster from the **Cluster** drop-down list.

Step 2 Click **More** and select **Download Client**. In the **Download Cluster Client** dialog box, select **Configuration Files Only**.

The generated compressed file contains the configuration files of all services.

Step 3 Determine whether to generate a configuration file on the cluster node.

- If yes, select **Save to Path**, and click **OK** to generate the client file. By default, the client file is generated in **/tmp/FusionInsight-Client** on the active management node. You can also store the client file in other directories, and user **omm** has the read, write, and execute permissions on the directories. Then, go to **Step 4**.
- If no, click **OK** and specify a local save path to download the complete client. Wait until the download is complete, and go to **Step 4**.

Step 4 Use WinSCP to save the compressed file to the installation directory of the client as the client installation user, such as **/opt/hadoopclient**.

Step 5 Decompress the software package.

Run the following commands to go to the directory where the client is installed, and decompress the file to a local directory. For example, the downloaded client file is **FusionInsight_Cluster_1_Services_Client.tar**.

cd /opt/hadoopclient

tar -xvf FusionInsight_Cluster_1_Services_Client.tar

Step 6 Verify the software package.

Run the following command to verify the decompressed file and check whether the command output is consistent with the information in the **sha256** file:

sha256sum -c FusionInsight_Cluster_1_Services_ClientConfig_ConfigFiles.tar.sha256

```
FusionInsight_Cluster_1_Services_ClientConfig_ConfigFiles.tar: OK
```

Step 7 Decompress the package to obtain the configuration file.

tar -xvf FusionInsight_Cluster_1_Services_ClientConfig_ConfigFiles.tar

Step 8 Run the following command in the client installation directory to update the client using the configuration file:

sh refreshConfig.sh *Client installation directory Directory where the configuration file is located*

For example, run the following command:

sh refreshConfig.sh /opt/hadoopclient /opt/hadoopclient/FusionInsight_Cluster_1_Services_ClientConfig_ConfigFiles

If the following information is displayed, the configurations have been updated successfully:

```
Succeed to refresh components client config.
```

----End

Method 2:

Step 1 Log in to the node where the client is installed as user **root**.

Step 2 Go to the client installation directory, for example, **/opt/client**, and run the following commands to update the configuration file:

cd /opt/client

sh autoRefreshConfig.sh

Step 3 Enter the username and password of the FusionInsight Manager administrator and the floating IP address of FusionInsight Manager.

Step 4 Enter the names of the components whose configurations need to be updated. Use commas (,) to separate the component names. Press **Enter** to update the configurations of all components if necessary.

If the following information is displayed, the configurations have been updated successfully:

```
Succeed to refresh components client config.
```

----End

9.9.2 Cluster Mutual Trust Management

9.9.2.1 Overview of Mutual Trust Between Clusters

Function Description

By default, users of a big data cluster in security mode can only access resources in the cluster but cannot perform identity authentication or access resources in other clusters in security mode.

Feature Description

- **Domain**

  The secure usage scope of users in each system is called a domain. Each FusionInsight Manager must have a unique domain name. Cross-Manager access allows users to use resources across domains.

- **User Encryption**

  Mutual trust can be configured across FusionInsight Managers. The current Kerberos server supports only the aes256-cts-hmac-sha1-96:normal and aes128-cts-hmac-sha1-96:normal encryption types for encrypting cross-domain users, and the encryption types cannot be changed.

- **User Authentication**

  After cross-Manager mutual trust is configured, if a user with the same name exists in two systems and the user in the peer system has the permission to access a resource in that system, this user can also access the remote resource.

- **Direct Mutual Trust**

  The system saves the mutual trust ticket of the peer system in two clusters with mutual trust configured and uses the mutual trust ticket to access the peer system.

9.9.2.2 Changing Manager's Domain Name

Scenario

The secure usage scope of users in each system is called a domain. Each system must have a unique domain name. The domain name of FusionInsight Manager is generated during installation. The system administrator can change the domain name on FusionInsight Manager.

NOTICE

- Changing the system domain name is a high-risk operation. Before performing operations in this section, ensure that the OMS data has been backed up by referring to **Backing Up Manager Data**.

Impact on the System

- During the configuration, all of the clusters need to be restarted and are unavailable during restart.
- After the domain name is changed, the passwords of the Kerberos administrator and OMS Kerberos administrator will be initialized. You need to use the default passwords and then change the passwords. If a component user whose password is generated randomly by the system is used for identity authentication, see **Exporting an Authentication Credential File** to download the keytab file again.
- After the domain name is changed, passwords of the **admin** user, component user, and human-machine user added by the system administrator before the domain name change will be reset to the same one. Change these passwords.

The reset password consists of two parts: one part is generated by the system and the other is set by the user. The system generating part is **Admin@123**, which is the default password. For details about the user-defined part, see descriptions of **Password Suffix** in **Table 9-54**. For example, if the system generates **Admin@123** and the user sets **Test#$%@123**, the new password after reset is **Admin@123Test#$%@123**.

- The new password must meet the password policies. To obtain the new human-machine user password, log in to the active OMS as user **omm** and run the following script:

  sh ${BIGDATA_HOME}/om-server/om/sbin/get_reset_pwd.sh *Password suffix user_name*

  - *Password suffix* is a parameter set by the user. If it is not specified, the default value **Admin@123** is used.
  - *user_name* is optional. The default value is **admin**.

  Example:

  sh ${BIGDATA_HOME}/om-server/om/sbin/get_reset_pwd.sh Test#$%@123

  ```
  To get the reset password after changing cluster domain name.
   pwd_min_len      : 8
   pwd_char_types   : 4
  The password reset after changing cluster domain name is: "Admin@123Test#$%@123"
  ```

  In this example, **pwd_min_len** and **pwd_char_types** indicate the minimum password length and number of password character types respectively defined in the password policies. **Admin@123Test#$%@123** indicates the human-machine user password after the system domain name is changed.

- After the system domain name is changed, the reset password consists of two parts: one part is generated by the system and the other is set by the user. The reset password must meet the password policies. If the password is not long enough, one or multiple at signs (@) are added between **Admin@123** and the user-defined part. If there are five character types, a space is added after **Admin@123**.

  When the user-defined part is **Test@123** and the default user password policy is used, the new password is **Admin@123Test@123**. The password contains 17 characters of four types. To meet the current password policy, the new password is processed according to **Table 9-53**.

Table 9-53 Password processing

Minimum Password Length	Number of Character Types	Processing Against the Password Policy	New Password
8 to 17 characters	4	The user password policy is met.	Admin@123Test@123
18 characters	4	Add an at sign (@).	Admin@123@Test@123
19 characters	4	Add two at signs (@).	Admin@123@@Test@123

Minimum Password Length	Number of Character Types	Processing Against the Password Policy	New Password
8 to 18 characters	5	Add a space.	Admin@123 Test@123
19 characters	5	Add a space and an at sign (@).	Admin@123 @Test@123
20 characters	5	Add a space and two at signs (@).	Admin@123 @@Test@123

- After the system domain name is changed, download the **keytab** file for the machine-machine user added by the system administrator before the domain name is changed.
- After the system domain name is changed, download and install the client again.

Prerequisites

- The system administrator has clarified service requirements and planned domain names for the systems.

  A domain name can contain only uppercase letters, numbers, periods (.), and underscores (_), and must start with a letter or number.
- The running status of all components in the Manager clusters is **Normal**.
- The **acl.compare.shortName** parameter of the ZooKeeper service of all clusters in Manager is set to default value **true**. Otherwise, change the value to **true** and restart the ZooKeeper service.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **Domain and Mutual Trust**.

Step 3 Modify required parameters.

Table 9-54 Related parameters

Parameter	Description
Local Domain	Planned domain name of the system.

Parameter	Description
Password Suffix	Part of the password set by the user after the password of the human-machine user is reset. This parameter is mandatory. The default value is **Admin@123**. **NOTE** This parameter takes effect only after **Local Domain** is modified. The following conditions must be met: ● The password ranges from 8 to 16 characters. ● The password must contain at least three types of the following: uppercase letters, lowercase letters, numbers, and special characters (`~!@#$%^&*()-_=+|[{}];:',<.>/? and spaces).

Step 4 Click **OK**. Proceed with the subsequent steps only after the modification is complete.

Step 5 Log in to the active management node as user **omm**.

Step 6 Run the following command to update the domain configuration:

```
sh ${BIGDATA_HOME}/om-server/om/sbin/restart-RealmConfig.sh
```

The command is executed successfully if the following information is displayed:

```
Modify realm successfully. Use the new password to log into FusionInsight again.
```

NOTE

After the restart, some hosts and services cannot be accessed and an alarm is generated. This problem can be automatically resolved in about 1 minute after **restart-RealmConfig.sh** is run.

Step 7 Log in to FusionInsight Manager using the new password of user **admin** (for example, **Admin@123Admin@123**). On the dashboard, click ··· next to the name of the target cluster and select **Restart**.

In the displayed dialog box, enter the password of the current login user and click **OK**.

In the displayed dialog box, click **OK**. Wait for a while until a message indicating that the operation is successful is displayed. Click **Finish**.

Step 8 Log out of FusionInsight Manager and then log in again. If the login is successful, the configuration is successful.

Step 9 Log in to the active management node as user **omm** and run the following command to update the configurations of the job submission client:

```
sh /opt/executor/bin/refresh-client-config.sh
```

Step 10 If a HetuEngine compute instance is running, restart the compute instance.

1. Log in to FusionInsight Manager as the user who is used to access the HetuEngine web UI.
2. Choose **Cluster** > **Services** > **HetuEngine** to go to the HetuEngine service page.
3. In the **Basic Information** area on the **Dashboard** page, click the link next to **HSConsole WebUI**. The HSConsole page is displayed.

4. For a running compute instance, click **Stop** in the **Operation** column. After the compute instance is in the **Stopped** state, click **Start** to restart the compute instance.

----End

9.9.2.3 Configuring Cross-Manager Mutual Trust Between Clusters

Scenario

When two security-mode clusters managed by different FusionInsight Managers need to access each other's resources, the system administrator can configure cross-Manager mutual trust for them.

The secure usage scope of users in each system is called a domain. Each FusionInsight Manager must have a unique domain name. Cross-Manager access allows users to use resources across domains.

NOTE

A maximum of 500 mutually trusted clusters can be configured.

Impact on the System

- After cross-Manager cluster mutual trust is configured, users of an external system can be used in the local system. The system administrator needs to periodically check the user permissions in Manager based on enterprise service and security requirements.
- When cross-Manager cluster mutual trust is configured, all clusters need to be stopped, causing service interruptions.
- After cross-Manager cluster mutual trust is configured, internal Kerberos users **krbtgt/***Local cluster domain name***@***External cluster domain name* and **krbtgt/***External cluster domain name***@***Local cluster domain name* are added to the two mutually trusted clusters. The internal users cannot be deleted. The system administrator needs to change the passwords periodically based on enterprise service and security requirements. The passwords of these four users in the two systems must be the same. For details, see **Changing the Password for a Component Running User**. When the passwords are changed, the connectivity between cross-cluster service applications may be affected.
- After cross-Manager cluster mutual trust is configured, the clients of each cluster need to be downloaded and installed again.
- After cross-Manager cluster mutual trust is configured, you need to check whether the system works properly and how to access resources of the peer system as a user of the local system. For details, see **Assigning User Permissions After Cross-Cluster Mutual Trust Is Configured**.

Prerequisites

- The system administrator has clarified service requirements and planned domain names for the systems. A domain name can contain only uppercase letters, numbers, periods (.), and underscores (_), and must start with a letter or number.

- The domain names of the two Managers are different. When an ECS or BMS cluster is created on MRS, a unique system domain name is randomly generated. Generally, you do not need to change the system domain name.
- The two clusters do not have the same host name or the same IP address.
- The system time of the two clusters is consistent, and the NTP services in the two systems use the same clock source.
- The running status of all components in the Manager clusters is **Normal**.
- The **acl.compare.shortName** parameter of the ZooKeeper service of all clusters in Manager is set to default value **true**. Otherwise, change the value to **true** and restart the ZooKeeper service.

Procedure

Step 1 Log in to one FusionInsight Manager.

Step 2 Stop all clusters on the home page.

Click ··· next to the target cluster and select **Stop**. Enter the password of the cluster administrator. In the **Stop Cluster** dialog box that is displayed, click **OK**. Wait until the cluster is stopped.

Step 3 Choose **System** > **Permission** > **Domain and Mutual Trust**.

Step 4 Modify **Peer Mutual Trust Domain**.

Table 9-55 Related parameters

Parameter	Description
realm_name	Enter the domain name of the peer system.

Parameter	Description
ip_port	Enter the KDC address of the peer system. Value format: *IP address of the node accommodating the Kerberos service in the peer system:Port number* ● In dual-plane networking, enter the service plane IP address. ● If an IPv6 address is used, the IP address must be enclosed in square brackets ([]). ● Use commas (,) to separate the KDC addresses if the active and standby Kerberos services are deployed or multiple clusters in the peer system need to establish mutual trust with the local system. ● You can obtain the port number from the **kdc_ports** parameter of the KrbServer service. The default value is **21732**. To obtain the IP address of the node where the service is deployed, click the **Instance** tab on the KrbServer page and view **Service IP Address** of the KerberosServer role. For example, if the Kerberos service is deployed on nodes at **10.0.0.1** and **10.0.0.2** that have established mutual trust with the local system, the parameter value is **10.0.0.1:21732,10.0.0.2:21732**.

NOTE

If you need to configure mutual trust for multiple Managers, click + to add a new item and set parameters. A maximum of 16 systems can be mutually trusted. Click − to delete unnecessary configurations.

Step 5 Click **OK**.

Step 6 Log in to the active management node as user **omm**, and run the following command to update the domain configuration:

sh ${BIGDATA_HOME}/om-server/om/sbin/restart-RealmConfig.sh

The command is executed successfully if the following information is displayed:

```
Modify realm successfully. Use the new password to log into FusionInsight again.
```

After the restart, some hosts and services cannot be accessed and an alarm is generated. This problem can be automatically resolved in about 1 minute after **restart-RealmConfig.sh** is run.

Step 7 Log in to FusionInsight Manager and start all clusters.

Click ⋯ next to the name of the target cluster and select **Start**. In the displayed **Start Cluster** dialog box, click **OK**. Wait until the cluster is started.

Step 8 Log in to the other FusionInsight Manager and repeat the preceding operations.

----End

9.9.2.4 Assigning User Permissions After Cross-Cluster Mutual Trust Is Configured

Scenario

After cross-Manager cluster mutual trust is configured, assign user access permissions on FusionInsight Managers so that these users can perform service operations in the mutually trusted Managers.

Prerequisites

The mutual trust between the two Managers has been configured.

Procedure

Step 1 Log in to the local FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User** to check whether the target user exists.

- If yes, go to **Step 3**.
- If no, go to **Step 4**.

Step 3 Click ⌄ on the left of the target user, and check whether the permissions assigned to the user group of the user and the roles meet service requirements. If not, create a role and bind the role to the user by referring to **Configuring Permissions**, or modify the user group or role permissions of the user.

Step 4 Create a user required by the service operations and associate the required user group or role. For details, see **Creating a User**.

Step 5 Log in to the other FusionInsight Manager and repeat **Step 2** to **Step 4** to create a user with the same name and set permissions.

----End

9.9.3 Configuring Scheduled Backup of Alarm and Audit Information

Scenario

You can modify the configuration file to periodically back up FusionInsight Manager alarm information, FusionInsight Manager audit information, and audit information of all services to the specified storage location.

The backup can be performed using FTP or SFTP. FTP does not encrypt data, which may cause security risks. Therefore, SFTP is recommended.

Procedure

Step 1 Log in to the active management node as user **omm**.

NOTE

Perform this operation only on the active management node. Scheduled backup is not supported on the standby management node.

Step 2 Run the following command to switch the directory:

cd ${BIGDATA_HOME}/om-server/om/sbin

Step 3 Run the following command to configure scheduled backup of FusionInsight Manager's alarm and audit information or service audit information:

./setNorthBound.sh -t *Information type* **-i** *Remote server IP address* **-p** *SFTP or FTP port used by the server***-u** *Username* **-d** *Save path* **-c** *Interval (minutes)* **-m** *Number of records in each file* **-s** *Whether to enable backup* **-e** *Protocol*

Example:

./setNorthBound.sh -t alarm -i 10.0.0.10 -p 22 -u sftpuser -d /tmp/ -c 10 -m 100 -s true -e sftp

This script modifies the alarm backup configuration file **alarm_collect_upload.properties**. The file save path is **${BIGDATA_HOME}/om-server/tomcat/webapps/web/WEB-INF/classes/config**.

./setNorthBound.sh -t audit -i 10.0.0.10 -p 22 -u sftpuser -d /tmp/ -c 10 -m 100 -s true -e sftp

This script modifies the audit backup configuration file **audit_collect_upload.properties**. The file save path is **${BIGDATA_HOME}/om-server/tomcat/webapps/web/WEB-INF/classes/config**.

./setNorthBound.sh -t service_audit -i 10.0.0.10 -p 22 -u sftpuser -d /tmp/ -c 10 -m 100 -s true -e sftp

This script modifies the service audit backup configuration file **service_audit_collect_upload.properties**. The file save path is **${BIGDATA_HOME}/om-server/tomcat/webapps/web/WEB-INF/classes/config**.

Step 4 Enter the password as prompted. The password is encrypted and saved in the configuration file.

```
Please input sftp/ftp server password:
```

Step 5 Check the configuration result. If the following information is displayed, the configuration is successful. The configuration file will be automatically synchronized to the standby management node.

```
execute command syncfile successfully.
Config Succeed.
```

----End

9.9.4 Modifying the FusionInsight Manager Routing Table

Scenario

When FusionInsight Manager is installed, two pieces of routing information are automatically created on the active management node. You can run the **ip rule list** command to view the routing information, as shown in the following example:

```
0:from all lookup local
32764:from all to 10.10.100.100 lookup ntp_rt   #NTP routing information created by FusionInsight
Manager (this information is unavailable if no external NTP clock source is configured).
32765:from 192.168.0.117 lookup om_rt   #OM routing information created by the FusionInsight Manager.
32766:from all lookup main
32767:from all lookup default
```

NOTE

If no external NTP server has been configured, only the OM routing information will be created.

If the routing information created by FusionInsight Manager conflicts with the routing information configured in the enterprise network planning, the cluster administrator can use **autoroute.sh** to disable or enable the routing information created by FusionInsight Manager.

Impact on the System

After the routing information created by FusionInsight Manager is disabled and before the new routing information is set, FusionInsight Manager cannot be accessed but the clusters are running properly.

Prerequisites

FusionInsight Manager has been installed.

You have obtained routing information about the WS floating IP address.

Disable the Routing Information Created by the System

Step 1 Log in to the active management node as user **omm**. Run the following commands to disable the routing information created by the system:

cd ${BIGDATA_HOME}/om-server/om/sbin

./autoroute.sh disable

```
Deactivating Route.
Route operation (disable) successful.
```

Step 2 Run the following command to view the execution result:

ip rule list

```
0:from all lookup local
32766:from all lookup main
32767:from all lookup default
```

Step 3 Run the following command and enter the password of user **root** to switch to user **root**:

su - root

Step 4 Run the following commands to manually create the routing information about the WS floating IP address:

ip route add *Network segment of the WS floating IP address/Subnet mask of the WS floating IP address* **scope link src** *WS floating IP address* **dev** *NIC of the WS floating IP address* **table om_rt**

ip route add default via *Gateway of the WS floating IP address* **dev** *NIC of the WS floating IP address* **table om_rt**

ip rule add from *WS floating IP address* **table om_rt**

Example:

ip route add 192.168.0.0/255.255.255.0 scope link src 192.168.0.117 dev eth0:ws table om_rt

ip route add default via 192.168.0.254 dev eth0:ws table om_rt

ip rule add from 192.168.0.117 table om_rt

NOTE

If IPv6 addresses are used, run the **ip -6 route add** command.

Step 5 Run the following commands to manually create the NTP service routing information. Skip this step when no external NTP clock source is configured.

ip route add default via *IP gateway of the NTP service* **dev** *NIC of the local IP address* **table ntp_rt**

ip rule add to *ntpIP* **table ntp_rt**

NIC of the local IP address indicates the NIC that can communicate with the network segment where the NTP server is located.

Example:

ip route add default via 10.10.100.254 dev eth0 table ntp_rt

ip rule add to 10.10.100.100 table ntp_rt

Step 6 View the execution result.

In the following example, if the command output contains **om_rt** and **ntp_rt**, the operation is successful.

ip rule list

```
0:from all lookup local
32764:from all to 10.10.100.100 lookup ntp_rt  #This information is not displayed if no external NTP clock source is configured.
32765:from 192.168.0.117 lookup om_rt
32766:from all lookup main
32767:from all lookup default
```

----End

Enable the Routing Information Created by the System

Step 1 Log in to the active management node as user **omm**.

Step 2 Run the following commands to enable the routing information created by the system:

cd ${BIGDATA_HOME}/om-server/om/sbin

./autoroute.sh enable

```
Activating Route.
Route operation (enable) successful.
```

Step 3 View the execution result.

In the following example, if the command output contains **om_rt** and **ntp_rt**, the operation is successful.

ip rule list

```
0:from all lookup local
32764:from all to 10.10.100.100 lookup ntp_rt  #This information is not displayed if no external NTP clock source is configured.
32765:from 192.168.0.117 lookup om_rt
32766:from all lookup main
32767:from all lookup default
```

----End

9.9.5 Switching to the Maintenance Mode

Scenario

FusionInsight Manager allows you to set clusters, services, hosts, or OMSs to the maintenance mode. Objects in maintenance mode do not report alarms. This prevents the system from generating a large number of unnecessary alarms during maintenance changes, such as upgrade, because these alarms may influence O&M personnel's judgment on the cluster status.

- Cluster maintenance mode

  If a cluster is not brought online or has been brought offline due to O&M operations (for example, non-rolling upgrade), you can set the entire cluster to the maintenance mode.

- Service maintenance mode

  When performing maintenance operations on a specific service (for example, performing service-affecting commissioning operations like batch restart of service instances, directly powering on or off nodes of the service, or repairing the service), you can set only this service to the maintenance mode.

- Host maintenance mode

  When performing maintenance operations on a host (such as powering on or off, isolating, or reinstalling the host, upgrading its OS, or replacing the host), you can set only this host to the maintenance mode.

- OMS maintenance mode

  When restarting, replacing, or repairing an OMS node, you can set the OMS node to the maintenance mode.

Impact on the System

After the maintenance mode is set, alarms caused by non-maintenance operations are suppressed and cannot be reported. Alarms can be reported only when faults persist after the system exits the maintenance mode. Therefore, exercise caution when setting the maintenance mode.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Set the maintenance mode.

Determine the object to set the maintenance mode based on the service scenario. For details, see **Table 9-56**.

Table 9-56 Setting to the maintenance mode

Scenario	Operation
Configure a cluster to enter the maintenance mode.	1. On FusionInsight Manager, click ··· next to the target cluster name and select **Enter Maintenance Mode**. 2. In the displayed dialog box, click **OK**. After the cluster enters the maintenance state, the status of the cluster becomes ⚙. After maintenance is complete, click **Exit Maintenance Mode**. The cluster then exits the maintenance mode.
Configure a service to enter the maintenance mode.	1. On FusionInsight Manager, choose **Cluster**, click the name of the desired cluster, choose **Services**, and click the service name. 2. On the service details page, click **More** and select **Enter Maintenance Mode**. 3. In the displayed dialog box, click **OK**. After a service enters the maintenance mode, the status of the service becomes ⚙ in the service list. After maintenance is complete, click **Exit Maintenance Mode**. The service then exits the maintenance mode. **NOTE** When configuring a service to enter the maintenance mode, you are advised to set the upper-layer services that depend on this service to the maintenance mode as well.
Configure a host to the maintenance mode.	1. On FusionInsight Manager, choose **Hosts**. 2. On the **Hosts** page, select the target host, click **More**, and select **Enter Maintenance Mode**. 3. In the displayed dialog box, click **OK**. After the host enters the maintenance mode, the status of the host becomes ⚙ in the host list. After maintenance is complete, click **Exit Maintenance Mode**. The host then exits the maintenance mode.
Configure the OMS to enter the maintenance mode.	1. On FusionInsight Manager, choose **System** > **OMS** > **Enter Maintenance Mode**. 2. In the displayed dialog box, click **OK**. After the OMS enters the maintenance state, the OMS status becomes ⚙. After maintenance is complete, click **Exit Maintenance Mode**. The OMS then exits the maintenance mode.

Step 3 Check the cluster maintenance view.

On FusionInsight Manager, click ··· next to the cluster name and select **Maintenance Mode View**. In the displayed window, you can view the services and hosts in maintenance mode in the cluster.

After maintenance is complete, you can select services and hosts in batches in the maintenance mode view and click **Exit Maintenance Mode** to make them exit the maintenance mode.

----End

9.9.6 Routine Maintenance

To ensure long-term and stable running of the system, administrators or maintenance engineers need to periodically check items listed in **Table 9-57** and rectify the detected faults based on the check results. It is recommended that administrators or engineers record the result in each task scenario and sign off based on the enterprise management regulations.

Table 9-57 Routine maintenance check items

Routine Maintenance Frequency	Task Scenario	Check Item
Daily	Check the cluster service status.	● Check whether the running status and configuration status of each service are normal and whether the status icons are green. ● Check whether the running status and configuration status of the role instances in each service are normal and whether the status icons are green. ● Check whether the active/standby status of role instances in each service can be properly displayed. ● Check whether the dashboard of the services and role instances can be displayed properly.
	Check the cluster host status.	● Check whether the running status of each host is normal and whether the status icon is green. ● Check the current disk usage, memory usage, and CPU usage of each host. Check whether the current memory usage and CPU usage are increasing.
	Check the cluster alarm information.	Check whether alarms were generated for unhandled exceptions on the previous day, including alarms that were automatically cleared.
	Check the cluster audit information.	Check whether critical and major operations are performed on the previous day and whether the operations are valid.

Routine Maintenance Frequency	Task Scenario	Check Item
	Check the cluster backup status.	Check whether OMS, DBService, NameNodeOMS, DBServiceOMS, and LDAP have been automatically backed up on the previous day.
	View the health check result.	Perform a health check on FusionInsight Manager and download the health check report to check whether the current cluster is abnormal. You are advised to enable the automatic health check, export the latest cluster health check result, and repair unhealthy items based on the result.
	Check the network communication.	Check the cluster network status and check whether the network communication between nodes is delayed.
	Check the storage status.	Check whether the total data storage volume of the cluster increases abruptly. ● Check whether the disk usage is close to the threshold. If yes, locate the causes. For example, check whether the junk data or cold data left by services needs to be cleared. ● Check whether disk partitions need to be expanded based on the service growth trend.
	Check logs.	● Check whether there are failed or unresponsive MapReduce and Spark tasks. Check the **/tmp/logs/${username}/logs/${application id}** log file in HDFS and rectify faults. ● Check Yarn task logs, view the logs of failed and unresponsive tasks, and delete duplicate data. ● Check the worker logs of Storm. ● Back up logs to the storage server.
Weekly	Manage users.	Check whether the user password is about to expire and notify the user of changing the password. To change the password of a machine-machine user, you need to download the keytab file again.
	Analyze alarms.	Export and analyze alarms generated in a specified period.

Routine Maintenance Frequency	Task Scenario	Check Item
	Scan disks.	Check the disk health status. You are advised to use a dedicated disk check tool.
	Collect statistics on storage.	Check in batches whether the disk data of cluster nodes is evenly stored, filter out the disks whose data increases significantly or is insufficient, and check whether the disks are normal.
	Record changes.	Arrange and record the operations on cluster configuration parameters and files to provide reference for fault analysis and handling.
Monthly	Analyze logs.	● Collect and analyze hardware logs of cluster node servers, such as BMC system logs. ● Collect and analyze the OS logs of the cluster node servers. ● Collect and analyze cluster logs.
	Diagnose the network.	Analyze the network health status of the cluster.
	Manage hardware.	Check the equipment room environment and clean the devices.

9.10 Log Management

9.10.1 About Logs

Log Description

MRS cluster logs are stored in the **/var/log/Bigdata** directory. The following table lists the log types.

Table 9-58 Log types

Log Type	Description
Installation logs	Installation logs record information about FusionInsight Manager, cluster, and service installation to help users locate installation errors.

Log Type	Description
Run logs	Run logs record the running track information, debugging information, status changes, potential problems, and error information generated during the running of services.
Audit logs	Audit logs record information about users' activities and operation instructions, which can be used to locate fault causes in security events and determine who are responsible for these faults.

The following table lists the MRS log directories.

Table 9-59 Log directories

Directory	Log
/var/log/Bigdata/audit	Component audit log.
/var/log/Bigdata/controller	Log collecting script log. Controller process log. Controller monitoring log.
/var/log/Bigdata/dbservice	DBService log.
/var/log/Bigdata/flume	Flume log.
/var/log/Bigdata/hbase	HBase log.
/var/log/Bigdata/hdfs	HDFS log.
/var/log/Bigdata/hive	Hive log.
/var/log/Bigdata/httpd	HTTPd log.
/var/log/Bigdata/hue	Hue log.
/var/log/Bigdata/kerberos	Kerberos log.
/var/log/Bigdata/ldapclient	LDAP client log.
/var/log/Bigdata/ldapserver	LDAP server log.
/var/log/Bigdata/loader	Loader log.
/var/log/Bigdata/logman	Logman script log management log.
/var/log/Bigdata/mapreduce	MapReduce log.
/var/log/Bigdata/nodeagent	NodeAgent log.
/var/log/Bigdata/okerberos	OMS Kerberos log.
/var/log/Bigdata/oldapserver	OMS LDAP log.
/var/log/Bigdata/metric_agent	Run log file of MetricAgent.

Directory	Log
/var/log/Bigdata/omm	**oms**: complex event processing log, alarm service log, HA log, authentication and authorization management log, and monitoring service run log of the OMM server. **oma**: installation log and run log of the OMM agent. **core**: dump log generated when the OMM agent and the HA process are suspended.
/var/log/Bigdata/spark2x	Spark2x log.
/var/log/Bigdata/sudo	Log generated when the **sudo** command is executed by user **omm**.
/var/log/Bigdata/timestamp	Time synchronization management log.
/var/log/Bigdata/tomcat	Tomcat log.
/var/log/Bigdata/watchdog	Watchdog log.
/var/log/Bigdata/yarn	Yarn log.
/var/log/Bigdata/zookeeper	ZooKeeper log.
/var/log/Bigdata/oozie	Oozie log.
/var/log/Bigdata/kafka	Kafka log.
/var/log/Bigdata/storm	Storm log.
/var/log/Bigdata/upgrade	OMS upgrade log.
/var/log/Bigdata/update-service	Upgrade service log.

NOTE

After the multi-instance function is enabled, if the system administrator adds multiple HBase, Hive, and Spark service instances, the log description, log level, and log format of the newly added service instances are the same as those of the original service logs. Service instance logs are stored separately in the **/var/log/Bigdata/***servicenameN* directory. The audit logs of the HBase and Hive service instances are stored in the **/var/log/Bigdata/audit/***servicenameN* directory. For example, the logs of HBase1 are stored in the **/var/log/Bigdata/hbase1** and **/var/log/Bigdata/audit/hbase1** directories.

Installation Logs

Table 9-60 Installation logs

Installation Log	Description
Configuration log	Records information about the configuration process before the installation.
FusionInsight Manager installation log	Records information about the two-node FusionInsight Manager installation.
Cluster installation log	Records information about the cluster installation.

Run Logs

Table 9-61 describes the running information recorded in run logs.

Table 9-61 Running information

Run Log	Description
Installation preparation log	Records information about preparations for the installation, such as the detection, configuration, and feedback operation information.
Process startup log	Records information about the commands executed during the process startup.
Process startup exception log	Records information about exceptions during process startup, such as dependent service errors and insufficient resources.
Process run log	Records information about the process running track information and debugging information, such as function entries and exits as well as cross-module interface messages.
Process running exception log	Records errors that cause process running errors, for example, the empty input objects or encoding or decoding failure.
Process running environment log	Records information about the process running environment, such as resource status and environment variables.
Script log	Records information about the script execution process.

Run Log	Description
Resource reclamation log	Records information about the resource reclaiming process.
Uninstallation clearing logs	Records information about operations performed during service uninstallation, such as directory and execution time deletion.

Audit Logs

Audit information recorded in audit logs includes FusionInsight Manager audit information and component audit information.

Table 9-62 Audit information of FusionInsight Manager

Operation Type	Operation
User management	Creating a user. Modifying a user. Deleting a user. Creating a user group. Modifying a user group. Deleting a group. Adding a role. Changing the user's roles. Deleting a role. Changing a password policy. Changing a password. Resetting a password. Logging in. Logging out. Unlocking the screen. Downloading the authentication credential. Unauthorized operation. Unlocking a user account. Locking a user account. Locking the screen. Exporting a user. Exporting a user group. Exporting a role.

Operation Type	Operation
Cluster management	Starting a cluster. Stopping a cluster. Restarting a cluster. Performing a rolling restart of a cluster. Restarting all expired instances. Saving configurations. Synchronizing cluster configurations. Customizing cluster monitoring metrics. Configuring monitoring dumping. Saving monitoring thresholds. Downloading a client configuration file. Configuring the northbound Syslog interface. Configuring the northbound SNMP interface. Clearing alarms using SNMP. Adding a trap target using SNMP. Deleting a trap target using SNMP. Checking alarms using SNMP. Synchronizing alarms using SNMP. Creating a threshold template. Deleting a threshold template. Applying a threshold template. Saving cluster monitoring configurations. Exporting configurations. Importing cluster configurations. Exporting an installation template. Modifying a threshold template. Canceling the application of a threshold template. Masking an alarm. Sending an alarm. Changing the OMS database password. Resetting the component database password. Restarting OMM and Controller. Starting the health check of a cluster. Importing a certificate file. Configuring SSO information. Deleting historical health check reports. Modifying cluster properties.

Operation Type	Operation
	Running maintenance commands in synchronous mode. Running maintenance commands in asynchronous mode. Customizing report monitoring metrics. Exporting report monitoring data. Runing a command in asynchronous mode using SNMP. Restarting the Web service. Customizing monitoring metrics for static resource pools. Exporting monitoring data of a static resource pool. Customizing dashboard monitoring metrics. Stopping a task. Restoring configurations. Modifying domain and mutual trust configurations. Modifying system parameters. Making a cluster enter the maintenance mode. Making a cluster exit the maintenance mode. Making OMS enter the maintenance mode. Making OMS exit the maintenance mode. Making services in a cluster exit the maintenance mode in batches. Modifying OMS configurations. Enabling threshold alarms. Synchronizing all cluster configurations.

Operation Type	Operation
Service management	Starting a service. Stopping a service. Synchronizing service configurations. Refreshing a service queue. Customizing service monitoring metrics. Restarting a service. Performing a rolling service restart. Exporting service monitoring data. Importing service configuration data. Starting the health check of a service. Configuring a service. Uploading a configuration file. Downloading a configuration file. Synchronizing instance configurations. Commissioning an instance. Decommissioning an instance. Starting an instance. Stopping an instance. Customizing instance monitoring metrics. Restarting an instance. Performing a rolling restart of an instance. Exporting instance monitoring data. Importing instance configuration data. Creating an instance group. Modifying an instance group. Deleting an instance group. Moving an instance to another instance group. Making a service enter the maintenance mode. Making a service exit the maintenance mode. Changing the name of a service. Modifying service association. Downloading monitoring data. Masking alarms. Unmasking alarms. Exporting report data of a service. Adding custom parameters for a report. Modifying custom parameters of a report. Deleting custom parameters of a report.

Operation Type	Operation
	Switching over control nodes. Adding a mount table. Modifying a mount table.
Host management	Setting a node rack. Starting all roles. Stopping all roles. Isolating a host. Canceling isolation of a host. Customizing host monitoring metrics. Exporting host monitoring data. Making a host enter the maintenance mode. Making a host exit the maintenance mode. Exporting basic host information. Exporting host distribution report data. Exporting host trend report data. Exporting host cluster report data. Exporting report data of a service. Customizing host cluster monitoring metrics. Customizing host cluster trend monitoring metrics.
Alarm management	Exporting alarms. Clearing alarms. Exporting events. Clearing alarms in batches.
Log collection	Collecting log files. Downloading log files. Collecting service stack information. Collecting instance stack information. Preparing service stack information. Preparing instance stack information. Clearing service stack information. Clearing instance stack information.
Audit log management	Modifying audit dumping configurations. Exporting audit logs.

Operation Type	Operation
Data backup and restoration	Creating a backup task. Executing a backup task. Executing backup tasks in batches. Stopping a backup task. Deleting a backup task. Modifying a backup task. Locking a backup task. Unlocking a backup task. Creating a restoration task. Executing a restoration task. Stopping a restoration task. Retrying a restoration task. Deleting a restoration task.

Operation Type	Operation
Multi-tenant management	Saving static configurations. Adding a tenant. Deleting a tenant. Associating a service with a tenant. Deleting a service from a tenant. Configuring resources. Creating a resource. Deleting a resource. Adding a resource pool. Modifying a resource pool. Deleting a resource pool. Restoring tenant data. Modifying global configurations of a tenant. Modifying queue configurations of a capacity scheduler. Modifying queue configurations of a super scheduler. Modifying resource distribution of a capacity scheduler. Clearing resource distribution of a capacity scheduler. Modifying resource distribution of a super scheduler. Clearing resource distribution of a super scheduler. Adding a resource catalog. Modifying a resource catalog. Deleting a resource catalog. Customizing tenant monitoring metrics.

Operation Type	Operation
Health check	Starting the health check of a cluster. Starting the health check of a service. Starting the health check of a host. Starting the health check of OMS. Starting the system health check. Updating the health check configurations. Exporting health check reports. Exporting health check results of a cluster. Exporting health check results of a service. Exporting health check results of a host. Deleting historical health check reports. Exporting historical health check reports. Downloading a health check report.

Table 9-63 Component audit information

Audit Log	Operation Type	Operation
ClickHouse audit log	Maintenance management	Granting permissions. Revoking permissions. Recording authentication and login information.
	Service operations	Creating databases or tables. Inserting, deleting, querying, and migrating data.
DBService audit log	Maintenance management	Performing backup restoration operations.
HBase audit log	Data definition language (DDL) statements	Creating a table. Deleting a table. Modifying a table. Adding a column family. Modifying a column family. Deleting a column family. Enabling a table. Disabling a table. Modifying user information. Changing a password. Logging in.

Audit Log	Operation Type	Operation
	Data manipulation language (DML) statements	Putting data (to the **hbase:meta**, **_ctmeta_**, and **hbase:acl** tables). Deleting data (from the **hbase:meta**, **_ctmeta_**, and **hbase:acl** tables). Checking and putting data (to the **hbase:meta**, **_ctmeta_**, and **hbase:acl** tables). Checking and deleting data (from the **hbase:meta**, **_ctmeta_**, and **hbase:acl** tables).
	Permission control	Assigning permissions to a user. Canceling permission assigning.
HDFS audit log	Permission management	Managing access permissions on files or folders. Managing the owner information of files or folders.
	File operations	Creating a folder. Creating a file. Opening a file. Appending file content. Changing a file name. Deleting a file or folder. Setting time property of a file. Setting the number of file copies. Merging files. Checking the file system. Linking to a file.
Hive audit log	Metadata operations	Defining metadata, such as creating databases and tables. Deleting metadata, such as deleting databases and tables. Modifying metadata, such as adding columns and renaming tables. Importing and exporting metadata.
	Data maintenance	Loading data to a table. Inserting data into a table.
	Permission management	Creating or deleting a role. Granting/Reclaiming roles. Granting/Reclaiming permissions.
Hue audit log	Service startup	Starting Hue.

Audit Log	Operation Type	Operation
	User operations	Logging in. Logging out.
	Task operations	Creating a task. Modifying a task. Deleting a task. Submitting a task. Saving a task. Updating the status of a task.
KrbServer audit log	Maintenance management	Changing the password of a Kerberos account. Adding a Kerberos account. Deleting a Kerberos account. Authenticating users.
LdapServer audit log	Maintenance management	Adding an OS user. Adding a user group. Adding a user to a user group. Deleting a user. Deleting a group.
Loader audit log	Security management	Logging in.
	Metadata management	Querying connector information. Querying a framework. Querying step information.
	Data source connection management	Querying a data source connection. Adding a data source connection. Updating a data source connection. Deleting a data source connection. Activating a data source connection. Disabling a data source connection.

Audit Log	Operation Type	Operation
	Job management	Querying a job. Creating a job. Updating a job. Deleting a job. Activating a job. Disabling a job. Querying all execution records of a job. Querying the latest execution record of a job. Submitting a job. Stopping a job.
MapReduce audit log	Application running	Starting a container request. Stopping a container request. After a container request is complete, the status of the request becomes successful. After a container request is complete, the status of the request becomes failed. After a container request is complete, the status of the request becomes suspended. Submitting a task. Ending a task.
Oozie audit log	Task management	Submitting a task. Starting a task. Killing a task. Suspending a task. Resuming a task. Running a task again.
Spark2x audit log	Metadata operations	Defining metadata, such as creating databases and tables. Deleting metadata, such as deleting databases and tables. Modifying metadata, such as adding columns and renaming tables. Importing and exporting metadata.
	Data maintenance	Loading data to a table. Inserting data into a table.

Audit Log	Operation Type	Operation
Storm audit log	Nimbus operations	Submitting a topology. Stopping a topology. Reallocating a topology. Deactivating a topology. Activating a topology.
	UI operations	Stopping a topology. Reallocating a topology. Deactivating a topology. Activating a topology.
Yarn audit log	Job submission	Submitting a job to a queue.
ZooKeeper audit log	Permission management	Setting access permissions to Znode.
	Znode operations	Creating Znodes. Deleting Znodes. Configuring Znode data.

FusionInsight Manager audit logs are stored in the database. You can view and export the audit logs on the **Audit** page.

The following table lists the directories to store component audit logs. Audit log files of some components are stored in **/var/log/Bigdata/audit**, such as HDFS, HBase, MapReduce, Hive, Hue, Yarn, Storm, and ZooKeeper. The component audit logs are automatically compressed and backed up to **/var/log/Bigdata/audit/bk** at 03: 00 every day. A maximum of latest 90 compressed backup files are retained, and the backup time cannot be changed. For details about how to configure the number of reserved audit log files, see **Configuring the Number of Local Audit Log Backups**.

Audit log files of other components are stored in the component log directory.

Table 9-64 Directories for storing component audit logs

Component	Audit Log Directory
DBService	/var/log/Bigdata/audit/dbservice/dbservice_audit.log

Component	Audit Log Directory
HBase	/var/log/Bigdata/audit/hbase/hm/hbase-audit-hmaster.log /var/log/Bigdata/audit/hbase/hm/hbase-ranger-audit-hmaster.log /var/log/Bigdata/audit/hbase/rs/hbase-audit-regionserver.log /var/log/Bigdata/audit/hbase/rs/hbase-ranger-audit-regionserver.log /var/log/Bigdata/audit/hbase/rt/hbase-audit-restserver.log /var/log/Bigdata/audit/hbase/ts/hbase-audit-thriftserver.log
HDFS	/var/log/Bigdata/audit/hdfs/nn/hdfs-audit-namenode.log /var/log/Bigdata/audit/hdfs/nn/ranger-plugin-audit.log /var/log/Bigdata/audit/hdfs/dn/hdfs-audit-datanode.log /var/log/Bigdata/audit/hdfs/jn/hdfs-audit-journalnode.log /var/log/Bigdata/audit/hdfs/zkfc/hdfs-audit-zkfc.log /var/log/Bigdata/audit/hdfs/httpfs/hdfs-audit-httpfs.log /var/log/Bigdata/audit/hdfs/router/hdfs-audit-router.log
Hive	/var/log/Bigdata/audit/hive/hiveserver/hive-audit.log /var/log/Bigdata/audit/hive/hiveserver/hive-rangeraudit.log /var/log/Bigdata/audit/hive/metastore/metastore-audit.log /var/log/Bigdata/audit/hive/webhcat/webhcat-audit.log
Hue	/var/log/Bigdata/audit/hue/hue-audits.log
Kafka	/var/log/Bigdata/audit/kafka/audit.log
Loader	/var/log/Bigdata/loader/audit/default.audit
MapReduce	/var/log/Bigdata/audit/mapreduce/jobhistory/mapred-audit-jobhistory.log
Oozie	/var/log/Bigdata/audit/oozie/oozie-audit.log
Spark2x	/var/log/Bigdata/audit/spark2x/jdbcserver/jdbcserver-audit.log /var/log/Bigdata/audit/spark2x/jdbcserver/ranger-audit.log /var/log/Bigdata/audit/spark2x/jobhistory/jobhistory-audit.log
Storm	/var/log/Bigdata/audit/storm/logviewer/audit.log /var/log/Bigdata/audit/storm/nimbus/audit.log /var/log/Bigdata/audit/storm/supervisor/audit.log /var/log/Bigdata/audit/storm/ui/audit.log
Yarn	/var/log/Bigdata/audit/yarn/rm/yarn-audit-resourcemanager.log /var/log/Bigdata/audit/yarn/rm/ranger-plugin-audit.log /var/log/Bigdata/audit/yarn/nm/yarn-audit-nodemanager.log

Component	Audit Log Directory
ZooKeeper	/var/log/Bigdata/audit/zookeeper/quorumpeer/zk-audit-quorumpeer.log

9.10.2 Manager Log List

Log Description

Log path: The default storage path of Manager log files is **/var/log/Bigdata/***Manager component*.

- ControllerService: **/var/log/Bigdata/controller/** (OMS installation and run logs)
- HTTPd: **/var/log/Bigdata/httpd** (HTTPd installation and run logs)
- Logman: **/var/log/Bigdata/logman** (log packaging tool logs)
- NodeAgent: **/var/log/Bigdata/nodeagent** (NodeAgent installation and run logs)
- okerberos: **/var/log/Bigdata/okerberos** (okerberos installation and run logs)
- oldapserver: **/var/log/Bigdata/oldapserver** (oldapserver installation and run logs)
- MetricAgent: **/var/log/Bigdata/metric_agent** (MetricAgent run logs)
- OMM: **/var/log/Bigdata/omm** (OMM installation and run logs)
- Timestamp: **/var/log/Bigdata/timestamp** (NodeAgent startup time logs)
- Tomcat: **/var/log/Bigdata/tomcat** (Web process logs)
- Watchdog: **/var/log/Bigdata/watchdog** (watchdog logs)
- Upgrade: **/var/log/Bigdata/upgrade** (OMS upgrade logs)
- UpdateService: **/var/log/Bigdata/update-service** (upgrade service logs)
- Sudo: **/var/log/Bigdata/sudo** (sudo script execution logs)
- OS: **/var/log/***message file* (OS system logs)
- OS performance: **/var/log/osperf** (OS performance statistics logs)
- OS statistics: **/var/log/osinfo/statistics** (OS parameter configuration logs)

Log archive rule:

The automatic compression and archiving function is enabled for Manager logs. By default, when the size of a log file exceeds 10 MB, the log file is automatically compressed. The naming rule of a compressed log file is as follows: <*Original log name*>-<*yyyy-mm-dd_hh-mm-ss*>.[*ID*].**log.zip** A maximum of 20 latest compressed files are retained.

Table 9-65 Manager logs

Log Type	Log File Name	Description
Controller run logs	controller.log	Log file that records component installation, upgrade, configuration, monitoring, alarm reporting, and routine O&M operations
	controller_client.log	Run log file of the Representational State Transfer (REST) APIs
	acs.log	ACS run log file
	acs_spnego.log	spnego user logs in ACS
	aos.log	AOS run log file
	plugin.log	AOS plug-in logs
	backupplugin.log	Run log file that records the backup and restoration operations
	controller_config.log	Configuration run log file
	controller_nodesetup.log	Controller loading task log file
	controller_root.log	System log file of the Controller process
	controller_trace.log	Log file that records the remote procedure call (RPC) communication between Controller and NodeAgent
	controller_monitor.log	Monitoring log file
	controller_fsm.log	State machine log file
	controller_alarm.log	Controller alarm log file
	controller_backup.log	Controller backup and recovery log file
	install.log, restore_package.log, installPack.log, distributeAdapterFiles.log, and install_os_optimization.log	OMS installation log file

Log Type	Log File Name	Description
	oms_ctl.log	OMS startup and stop log file
	preInstall_client.log	Preprocessing log file before client installation
	installntp.log	NTP installation log file
	modify_manager_param.log	Manager parameter modification log file
	backup.log	OMS backup script run log file
	supressionAlarm.log	Alarm script run log file
	om.log	OM certificate generation log file
	backupplugin_ctl.log	Startup log file of the backup and restoration plug-in process
	getLogs.log	Run log of the log collection script
	backupAuditLogs.log	Run log of the audit log backup script
	certStatus.log	Log file that records regular certificate checks
	distribute.log	Certificate distribution log
	ficertgenetrate.log	Certificate replacement log file, covering level-2 certificates, CAS certificates, and HTTPd certificates
	genPwFile.log	Log file that records the generation of certificate password files
	modifyproxyconf.log	Log file that records the modification of the HTTPd proxy configuration
	importTar.log	Log file that records the process for importing certificates into the trust store.

Log Type	Log File Name	Description
HTTPd	install.log	HTTPd installation log file
	access_log, error_log	HTTPd run log file
Logman	logman.log	Log packaging tool log file
NodeAgent	install.log and install_os_optimization.log	NodeAgent installation log file
	installntp.log	NTP installation log file
	start_ntp.log	NTP startup log file
	ntpChecker.log	NTP check log file
	ntpMonitor.log	NTP monitoring log file
	heartbeat_trace.log	Log file that records heartbeats between NodeAgent and Controller
	alarm.log	Alarm log file
	monitor.log	Monitoring log file
	nodeagent_ctl.log and start-agent.log	NodeAgent startup log file
	agent.log	NodeAgent run log file
	cert.log	Certificate log file
	agentplugin.log	Log file that records the Agent plug-in running status
	omaplugin.log	OMA plug-in run log file
	diskhealth.log	Disk health check log file
	supressionAlarm.log	Alarm script run log file
	updateHostFile.log	Host list update log file
	collectLog.log	Run log file of the node log collection script
	host_metric_collect.log	Run log file of host metric collection

Log Type	Log File Name	Description
	checkfileconfig.log	Run log file of file permission check
	entropycheck.log	Entropy check run log file
	timer.log	Log file of scheduled node scheduling
	pluginmonitor.log	Component monitoring plug-in log file
	agent_alarm_py.log	Log file that records alarms upon insufficient NodeAgent file permission
oKerberos	addRealm.log and modifyKerberosRealm.log	Realm handover log file
	checkservice_detail.log	Okerberos health check log file
	genKeytab.log	keytab generation log file
	KerberosAdmin_genConfigDetail.log	Run log file of **kadmin.conf** generated during start of the kadmin process
	KerberosServer_genConfigDetail.log	Run log file of **krb5kdc.conf** generated during start of the krb5kdc process
	oms-kadmind.log	Run log file of the kadmin process
	oms_kerberos_install.log and postinstall_detail.log	Okerberos installation log file
	oms-krb5kdc.log	Run log file of the krbkdc process
	start_detail.log	Okerberos startup log file
	realmDataConfigProcess.log	Log file that records the rollback upon a realm handover failure
	stop_detail.log	Okerberos stop log file

Log Type	Log File Name	Description
oldapserver	ldapserver_backup.log	Oldapserver backup log file
	ldapserver_chk_service.log	Oldapserver health check log file
	ldapserver_install.log	Oldapserver installation log file
	ldapserver_start.log	Oldapserver startup log file
	ldapserver_status.log	Log file that records the status of the Oldapserver process
	ldapserver_stop.log	Oldapserver stop log file
	ldapserver_wrap.log	Oldapserver service management log file
	ldapserver_uninstall.log	Oldapserver uninstallation log file
	restart_service.log	Oldapserver restart log file
	ldapserver_unlockUser.log	Log file that records information about unlocking LDAP users and managing accounts
metric_agent	gc.log	MetricAgent JVM GC log file
	metric_agent.log	Run log file of MetricAgent
	metric_agent_qps.log	Log file that records MetricAgent Internal queue length and QPS information
	metric_agent_root.log	All run log files of MetricAgent
	start.log	Log file that records information about the MetricAgent startup and stop
OMM	omsconfig.log	OMS configuration log file
	check_oms_heartbeat.log	OMS heartbeat log file

Log Type	Log File Name	Description
	monitor.log	OMS monitoring log file
	ha_monitor.log	HA_Monitor operation log file
	ha.log	HA operation log file
	fms.log	Alarm log file
	fms_ha.log	HA alarm monitoring log file
	fms_script.log	Alarm control log file
	config.log	Alarm configuration log file
	iam.log	IAM log file
	iam_script.log	IAM control log file
	iam_ha.log	IAM HA monitoring log file
	config.log	IAM configuration log file
	operatelog.log	IAM operation log file
	heartbeatcheck_ha.log	OMS heartbeat HA monitoring log file
	install_oms.log	OMS installation log file
	pms_ha.log	HA monitoring log file
	pms_script.log	Monitoring control log file
	config.log	Monitoring configuration log file
	plugin.log	Monitoring plug-in run log file
	pms.log	Monitoring log file
	ha.log	HA run log file
	cep_ha.log	CEP HA monitoring log file
	cep_script.log	CEP control log file
	cep.log	CEP log file

Log Type	Log File Name	Description
	config.log	CEP configuration log file
	omm_gaussdba.log	GaussDB HA monitoring log file
	gaussdb-<SERIAL>.log	GaussDB run log file
	gs_ctl-<DATE>.log	Archive log file of GaussDB control logs
	gs_ctl-current.log	GaussDB control log file
	gs_guc-current.log	GaussDB operation log file
	encrypt.log	OMM encryption log file
	omm_agent_ctl.log	OMA control log file
	oma_monitor.log	OMA monitoring log file
	install_oma.log	OMA installation log file
	config_oma.log	OMA configuration log file
	omm_agent.log	OMA run log file
	acs.log	ACS resource log file
	aos.log	AOS resource log file
	controller.log	Controller resource log file
	floatip.log	Floating IP address resource log file
	ha_ntp.log	NTP resource log file
	httpd.log	HTTPd resource log file
	okerberos.log	Okerberos resource log file
	oldap.log	OLdap resource log file
	tomcat.log	Tomcat resource log file
	send_alarm.log	Run log file of the HA alarm sending script of the management node

Log Type	Log File Name	Description
	feed_watchdog.log	feed_watchdog resource log
Timestamp	restart_stamp	NodeAgent startup time log file
Tomcat	cas.log and localhost_access_cas_log.log	CAS run log file
	catalina.log, catalina.out, host-manager.log, localhost.log, and manager.log	Tomcat run log file
	localhost_access_web_log.log	Log file that records the access to REST APIs of FusionInsight Manager
	web.log	Run log file of the Web process
	northbound_ftp_sftp.log and snmp.log	Northbound log file
	perfStats.log	Performance statistics log file
Watchdog	watchdog.log and feed_watchdog.log	watchdog.log run log file
update-service	omm_upd_server.log	UPDServer run log file
	omm_upd_agent.log	UPDAgent run log file
	update-manager.log	UPDManager run log file
	install.log	Installation log file of the upgrade service
	uninstall.log	Uninstallation log file of the upgrade service

Log Type	Log File Name	Description
	catalina.<*Time*>.log, catalina.out, host-manager.<*Time*>.log, localhost.<*Time*>.log, manager.<*Time*>.log, manager_access_log.<*Time*>.txt, web_service_access_log.<*Time*>.txt, catalina.log, gc-update-service.log.0.current, update-manager.controller, update-web-service.controller, update-web-service.log, commit_rm_distributed.log, commit_rm_upload_package.log, common_omagent_operator.log, forbid_monitor.log, initialize_package_atoms.log, initialize_unzip_pack.log, omm-upd.log, register_patch_pack.log, resume_monitor.logrollback_clear_patch.log, unregister_patch_pack.log, update-rcommupd.log, update-rcupdatemanager.log, and update-service.log	Run log file of the upgrade service
Upgrade	upgrade.log_<*Time*>	OMS upgrade log file
	rollback.log_<*Time*>	OMS rollback log file
sudo	sudo.log	Sudo script execution log file

Log Levels

Table 9-66 describes the log levels provided by Manager. The log levels are FATAL, ERROR, WARN, INFO, and DEBUG in descending order. Logs whose levels are higher than or equal to the set level are printed by the program. The number of printed logs decreases as the set log level increases.

Table 9-66 Log levels

Level	Description
FATAL	Logs of this level record fatal error information about the current event processing that may result in a system crash.
ERROR	Logs of this level record error information about the current event processing, which indicates that system running is abnormal.
WARN	Logs of this level record abnormal information about the current event processing. These abnormalities will not result in system faults.
INFO	Logs of this level record normal running status information about the system and events.
DEBUG	Logs of this level record system information and debugging information.

Log Formats

The following table lists the Manager log formats.

Table 9-67 Log formats

Log Type	Component	Format	Example
Controller, HTTPd, Logman, NodeAgent, oKerberos, oldapserver, OMM, Tomcat, and upgrade	Controller, HTTPd, Logman, NodeAgent, oKerberos, oldapserver, OMM, Tomcat, and upgrade	<yyyy-MM-dd HH:mm:ss, SSS>|<Log Level>|<*Name of the thread for which the log is generated*>|<*Log message*>|<*Location where the log event occurs*>	2020-06-30 00:37:09,067 INFO [pool-1-thread-1] Completed Discovering Node. com.xxx.hadoop.om.controller.tasks.nodesetup.DiscoverNodeTask.execute(DiscoverNodeTask.java:299)

9.10.3 Configuring the Log Level and Log File Size

Scenario

You can change the log levels of FusionInsight Manager. For a specific service, you can change the log level and the log file size to prevent the failure in saving logs due to insufficient disk space.

Impact on the System

The services need to be restarted for the new configuration to take effect. During the restart, the services are unavailable.

Changing the FusionInsight Manager Log Level

1. Log in to the active management node as user **omm**.
2. Run the following command to switch to the required directory:

   cd ${BIGDATA_HOME}/om-server/om/sbin
3. Run the following command to change the log level:

   ./setLogLevel.sh*Log level parameters*

   The priorities of log levels are FATAL, ERROR, WARN, INFO, and DEBUG in descending order. Logs whose levels are higher than or equal to the set level are printed. The number of printed logs decreases as the configured log level increases.
   - **DEFAULT**: After this parameter is set, the default log level is used.
   - **FATAL**: critical error log level. After this parameter is set, only logs of the **FATAL** level are printed.
   - **ERROR**: error log level. After this parameter is set, logs of the **ERROR** and **FATAL** levels are printed.
   - **WARN**: warning log level. After this parameter is set, logs of the **WARN**, **ERROR**, and **FATAL** levels are printed.
   - **INFO** (default): informational log level. After this parameter is set, logs of the **INFO**, **WARN**, **ERROR**, and **FATAL** levels are printed.
   - **DEBUG**: debugging log level. After this parameter is set, logs of the **DEBUG**, **INFO**, **WARN**, **ERROR**, and **FATAL** levels are printed.
   - **TRACE**: tracing log level. After this parameter is set, logs of the **TRACE**, **DEBUG**, **INFO**, **WARN**, **ERROR**, and **FATAL** levels are printed.

   NOTE

   The log levels of components are different from those defined in open-source code.
4. Download and view logs to verify that the log level settings have taken effect. For details, see **Log**.

Changing the Service Log Level and Log File Size

NOTE

KrbServer, LdapServer, and DBService do not support the changing of service log levels and log file sizes.

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 3 Click a service in the service list. On the displayed page, click the **Configuration** page.

Step 4 On the displayed page, click the **All Configuration** tab. Expand the role instance displayed on the left of the page. Click **Log** of the role to be modified.

Step 5 Search for each parameter and obtain the parameter description. On the parameter configuration page, select the required log level or change the log file size. The unit of the log file size is MB.

NOTICE

- The system automatically deletes logs based on the configured log size. To save more information, set the log file size to a larger value. To ensure the integrity of log files, you are advised to manually back up the log files to another directory based on the actual service volume before the log files are cleared according to clearance rules.
- Some services do not support change of the log level on the UI.

Step 6 Click **Save**. In the **Save Configuration** dialog box, click **OK**.

Step 7 Download and view logs to verify that the log level settings have taken effect.

----End

9.10.4 Configuring the Number of Local Audit Log Backups

Scenario

Audit logs of cluster components are classified by name and stored in the **/var/log/Bigdata/audit** directory on each cluster node. The OMS automatically backs up the audit log directories at 03:00 every day.

The audit log directory on each node is compressed and named in the *<Node IP address>***.tar.gz** format. All compressed files are compressed and named in the *<yyyy-MM-dd_HH-mm-ss>***.tar.gz** format and saved in the **/var/log/Bigdata/audit/bk/** directory on the active management node. In addition, the standby management node saves a copy of the file.

By default, a maximum of 90 OMS backup files can be retained. This section describes how to configure the maximum number.

Procedure

Step 1 Log in to the active management node as user **omm**.

NOTE

Perform this operation only on the active management node. This operation is not supported on the standby management nodes; otherwise, the cluster cannot work properly.

Step 2 Run the following command to switch to the required directory:

cd ${BIGDATA_HOME}/om-server/om/sbin

Step 3 Run the following command to change the maximum number of audit log backup files to be retained:

./modifyLogConfig.sh -m *Maximum number of backup files that can be retained*

The default value is **90**. The value ranges from **0** to **365**. A larger value means to consume more disk space.

If the following information is displayed, the operation is successful:

```
Modify log config successfully
```

----End

9.10.5 Viewing Role Instance Logs

Scenario

FusionInsight Manager allows users to view logs of each role instance.

Procedure

Step 1 Log in to FusionInsight Manager.

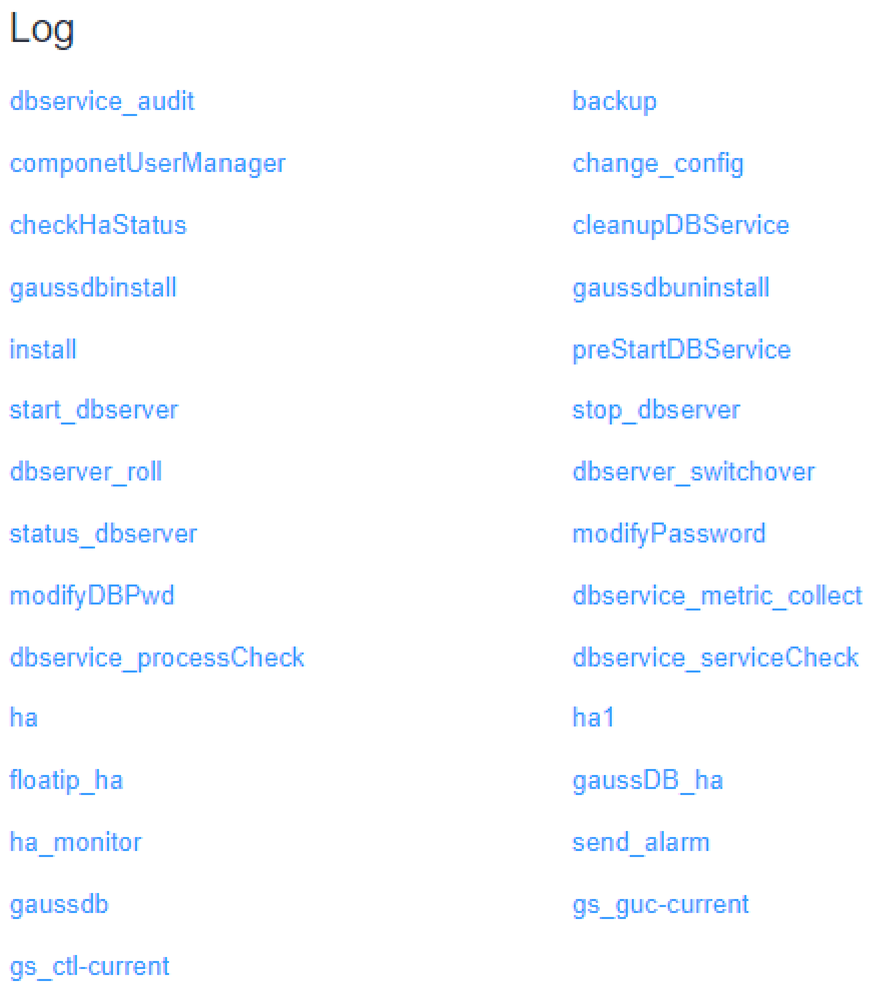
Step 2 Choose **Cluster**, click the name of the desired cluster, choose **Services**, and click a service name. Then click the **Instance** tab of the service and click the name of the target instance to access the instance status page.

Step 3 In the **Log** area, click the name of a log file to preview its content online.

NOTE

- On the **Hosts** page, click a host name. In the instance list of the host, you can view the log files of all role instances on the host.
- By default, a maximum of 100 lines of logs can be displayed. You can click **Load More** to view more logs. Click **Download** to download the log file to the local PC. For details about how to download service logs in batches, see **Log Download**.

Figure 9-13 Viewing instance logs



Log

dbservice_audit
backup
componetUserManager
change_config
checkHaStatus
cleanupDBService
gaussdbinstall
gaussdbuninstall
install
preStartDBService
start_dbserver
stop_dbserver
dbserver_roll
dbserver_switchover
status_dbserver
modifyPassword
modifyDBPwd
dbservice_metric_collect
dbservice_processCheck
dbservice_serviceCheck
ha
ha1
floatip_ha
gaussDB_ha
ha_monitor
send_alarm
gaussdb
gs_guc-current
gs_ctl-current

----End

9.11 Backup and Recovery Management

9.11.1 Introduction

Overview

FusionInsight Manager provides the backup and restoration of system data and user data by component. The system can back up Manager data, component metadata, and service data.

Data can be backed up to local disks (LocalDir), local HDFS (LocalHDFS), remote HDFS (RemoteHDFS), NAS (NFS/CIFS), Object Storage Service (OBS), and SFTP server (SFTP). For details, see **Backing Up Data**.

For a component that supports multiple services, multiple instances of a service can be backed up and restored. The backup and restoration operations are consistent with those of a service instance.

NOTE

Only MRS 3.1.0 or later supports data backup to OBS.

Backup and restoration tasks are performed in the following scenarios:

- Routine backup is performed to ensure the data security of the system and components.
- If the system is faulty, the data backup can be used to recover the system.
- If the active cluster is completely faulty, a mirrored cluster identical to the active cluster needs to be created. You can use the backup data to restore the active cluster.

Table 9-68 Manager configuration data to be backed up

Backup Type	Backup Content	Backup Directory Type
OMS	Database data (excluding alarm data) and configuration data in the cluster management system by default	● LocalDir ● LocalHDFS ● RemoteHDFS ● NFS ● CIFS ● SFTP ● OBS

Table 9-69 Component metadata or other data to be backed up

Backup Type	Backup Content	Backup Directory Type
DBService	Metadata of the components (including Loader, Hive, Spark, Oozie, and Hue) managed by DBService. For a cluster with multiple services installed, back up the metadata of multiple Hive and Spark service instances.	● LocalDir ● LocalHDFS ● RemoteHDFS ● NFS ● CIFS ● SFTP ● OBS
Kafka	Kafka metadata.	● LocalDir ● LocalHDFS ● RemoteHDFS ● NFS ● CIFS ● OBS

Backup Type	Backup Content	Backup Directory Type
NameNode	HDFS metadata. After multiple NameServices are added, backup and restoration are supported for all of them and the operations are consistent with those of the default hacluster instance.	- LocalDir - RemoteHDFS - NFS - CIFS - SFTP - OBS
Yarn	Information about the Yarn service resource pool.	
HBase	**tableinfo** files and data files of HBase system tables.	

Table 9-70 Service data of specific components to be backed up

Backup Type	Backup Content	Backup Directory Type
HBase	Table-level user data. For a cluster with multiple services installed, backup and restoration are supported for multiple HBase service instances and the backup and restoration operations are consistent with those of a single HBase service instance.	- RemoteHDFS - NFS - CIFS - SFTP
HDFS	Directories or files of user services. **NOTE** Encrypted directories cannot be backed up or restored.	
Hive	Table-level user data. For a cluster with multiple services installed, backup and restoration are supported for multiple Hive service instances and the backup and restoration operations are consistent with those of a single Hive service instance.	

Note that some components do not provide data backup or restoration:

- Kafka supports replicas and allows multiple replicas to be specified when a topic is created.
- MapReduce and Yarn data is stored in HDFS. Therefore, they rely on the backup and restoration provided by HDFS.
- Backup and restoration of service data in ZooKeeper are performed by their own upper-layer components.

Principles

Task

Before backup or restoration, you need to create a backup or restoration task and set task parameters, such as the task name, backup data source, and type of the directory for storing backup files. Then you can execute the tasks to back up or restore data. When Manager is used to restore the data of HDFS, HBase, Hive, and NameNode, the cluster cannot be accessed.

Each backup task can back up data of different data sources and generate an independent backup file for each data source. All the backup files generated in a backup task form a backup file set, which can be used in restoration tasks. Backup data can be stored on Linux local disks, local cluster HDFS, and standby cluster HDFS.

Backup tasks support full backup and incremental backup policies. Cloud data backup tasks do not support incremental backup. If the backup directory type is NFS or CIFS, incremental backup is not recommended. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

NOTE

Task execution rules:

- If a task is being executed, the task cannot be executed repeatedly and other tasks cannot be started at the same time.
- The interval at which a periodic task is automatically executed must be greater than 120s. Otherwise, the task is postponed and will be executed in the next period. Manual tasks can be executed at any interval.
- When a periodic task is to be automatically executed, the current time cannot be 120s later than the task start time. Otherwise, the task is postponed and executed in the next period.
- When a periodic task is locked, it cannot be automatically executed and needs to be manually unlocked.
- Before an OMS, DBService, Kafka, or NameNode backup task starts, ensure that the LocalBackup partition on the active management node has not less than 20 GB of available space. Otherwise, the backup task cannot be started.

When planning backup and restoration tasks, select the data to be backed up or restored strictly based on the service logic, data store structure, and database or table association. By default, the system creates periodic backup tasks **default-oms** and **default-*cluster ID*** at an interval of one hour. OMS metadata and cluster metadata, such as DBService and NameNode, can be fully backed up to local disks.

Snapshot

The system uses the snapshot technology to quickly back up data. Snapshots include HBase and HDFS snapshots.

- HBase snapshots

  An HBase snapshot is a backup file of HBase tables at a specified time point. This backup file does not replicate service data or affect the RegionServer. The HBase snapshot replicates table metadata, including table descriptor, region info, and HFile reference information. The metadata can be used to restore data before the snapshot creation time.

- HDFS snapshots

  An HDFS snapshot is a read-only backup of HDFS at a specified time point. The snapshot is used in data backup, misoperation protection, and disaster recovery scenarios.

  The snapshot function can be enabled for any HDFS directory to create the related snapshot file. Before creating a snapshot for a directory, the system automatically enables the snapshot function for the directory. Creating a snapshot does not affect any HDFS operation. A maximum of 65,536 snapshots can be created for each HDFS directory.

  When a snapshot is being created for an HDFS directory, the directory cannot be deleted or modified before the snapshot is created. Snapshots cannot be created for the upper-layer directories or subdirectories of the directory.

DistCp

Distributed copy (DistCp) is a tool used to replicate a large amount of data in HDFS in a cluster or between the HDFSs of different clusters. In a backup or restoration task of HBase, HDFS, or Hive, if you back up the data to HDFS of the standby cluster, the system invokes DistCp to perform the operation. Install the MRS software of the same version for the active and standby clusters and install the cluster.

DistCp uses MapReduce to implement data distribution, troubleshooting, restoration, and report. DistCp specifies different Map jobs for various source files and directories in the specified list. Each Map job copies the data in the partition that corresponds to the specified file in the list.

If you use DistCp to replicate data between HDFSs of two clusters, configure the cross-cluster mutual trust (mutual trust does not need to be configured for clusters managed by the same FusionInsight Manager) and cross-cluster replication for both clusters. When backing up the cluster data to HDFS in another cluster, you need to install the Yarn component. Otherwise, the backup fails.

Local rapid restoration

After using DistCp to back up the HBase, HDFS, and Hive data of the local cluster to the HDFS of the standby cluster, the HDFS of the local cluster retains the backup data snapshots. You can create local rapid restoration tasks to restore data by using the snapshot files in the HDFS of the local cluster.

NAS

Network Attached Storage (NAS) is a dedicated data storage server which includes the storage components and embedded system software. It provides the cross-platform file sharing function. By using NFS (supporting NFSv3 and NFSv4) and CIFS (supporting SMBv2 and SMBv3), you can connect the service plane of MRS to the NAS server to back up data to the NAS or restore data from the NAS.

NOTE

- Before data is backed up to the NAS, the system automatically mounts the NAS shared address to a local partition of the backup task execution node. After the backup is complete, the system unmounts the NAS shared partition from the backup task execution node.
- To prevent backup and restoration failures, do not access the shared address where the NAS server has been mounted to, for example, **/srv/BigData/LocalBackup/nas**, during data backup and restoration.
- When service data is backed up to the NAS, DistCp is used.

Specifications

Table 9-71 Specifications of the backup and restoration feature

Item	Specification
Maximum number of backup or restoration tasks	100
Number of concurrent tasks in a cluster	1
Maximum number of waiting tasks	199
Maximum size (GB) of backup files on a Linux local disk	600

NOTE

If service data is stored in the ZooKeeper upper-layer components, ensure that the number of znodes in a single backup or restoration task is not too large. Otherwise, the task will fail, and the ZooKeeper service performance will be affected. To check the number of znodes in a single backup or restoration task, perform the following operations:

- Ensure that the number of znodes in a single backup or restoration task is smaller than the upper limit of OS file handles. Specifically:
  1. To check the upper limit at the system level, run the **cat /proc/sys/fs/file-max** command.
  2. To check the upper limit at the user level, run the **ulimit -n** command.
- If the number of znodes in the parent directory exceeds the upper limit, back up and restore data in its sub-directories in batches. To check the number of znodes using ZooKeeper client scripts, perform the following operations:
  1. On FusionInsight Manager, choose **Cluster**, click the name of the desired cluster, choose **Services** > **ZooKeeper** > **Instance**, and view the management IP address of each ZooKeeper role.
  2. Log in to the node where the client is located and run the following command:

     zkCli.sh -server *ip*:*port*, where, *ip* can be any management IP address, and the default port number is 2181.
  3. If the following information is displayed, login to the ZooKeeper server is successful:
     ```
     WatchedEvent state:SyncConnected type:None path:null
     [zk: ip:port(CONNECIED) 0]
     ```
  4. Run the **getusage** command to check the number of znodes in the directory to be backed up.

     For example, **getusage /hbase/region**. In the command output, **Node count=xxxxxx** indicates the number of znodes stored in the **region** directory.

Table 9-72 Specifications of the default task

Item	OMS	HBase	Kafka	DBService	NameNode
Backup period	1 hour				
Maximum number of backups	168 (7-day historical data)				24 (one-day historical data)
Maximum size of a backup file	10 MB	10 MB	512 MB	100 MB	20 GB
Maximum size of disk space used	1.64 GB	1.64 GB	84 GB	16.41 GB	480 GB
Storage path of backup data	*Data storage path*/**LocalBackup/** of the active and standby management nodes				

NOTE

- The backup data of the default backup task must be periodically transferred and saved outside the cluster based on the enterprise O&M requirements.
- Administrators can create DistCp backup tasks to save OMS, DBService, and NameNode data to external clusters.
- The execution time of a cluster data backup task can be calculated using the following formula: Task execution time = Volume of data to be backed up/Network bandwidth between the cluster and the backup device. In practice, you are advised to multiply the calculated time by 1.5 to get the reference value of the task execution time.
- Executing a data backup task affects the maximum I/O performance of the cluster. Therefore, you are advised to execute a backup task during off-peak hours.

9.11.2 Backing Up Data

9.11.2.1 Backing Up Manager Data

Scenario

To ensure data security of FusionInsight Manager routinely or before and after a critical operation (such as capacity expansion and reduction) on FusionInsight Manager, you need to back up FusionInsight Manager data. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up Manager data. Both automatic and manual backup tasks are supported.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path***/LocalBackup/** has sufficient space on the active and standby management nodes.
- If you want to back up data to NAS, you have deployed the NAS server in advance.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Set **Backup Object** to **OMS**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 9-73 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	● **Full backup at the first time and incremental backup subsequently** ● **Full backup every time** ● **Full backup once every n times** **NOTE** ● Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. ● If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, select **OMS**.

Step 7 Set **Path Type** of **OMS** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files.

  The default storage directory is *Data storage path*/**LocalBackup/**, for example, **/srv/BigData/LocalBackup**.

  If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

If you select this option, set the following parameters:

  - **Target Path**: indicates the HDFS directory for storing the backup files. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Cluster for Backup**: Enter the cluster name mapping to the backup directory.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  - **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Source Cluster**: Select the cluster of the Yarn queue used by the backup data.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the source cluster.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select this option, set the following parameters:

  – **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.

  – **Server IP Address**: indicates the IP address of the NAS server.

  – **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.

  – **Username**: indicates the username set when the CIFS protocol is configured.

  – **Password**: indicates the password set when the CIFS protocol is configured.

  – **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)

  – **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:

  – **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.

  – **Server IP Address**: indicates the IP address of the server where the backup data is stored.

  – **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.

  – **Username**: indicates the username for connecting to the server using the SFTP protocol.

  – **Password**: indicates the password for connecting to the server using the SFTP protocol.

  – **Server Shared Path**: indicates the backup path on the SFTP server.

  – **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **OBS**: indicates that backup files are stored in OBS.

  If you select this option, set the following parameters:

  – **Target Path**: indicates the OBS directory for storing backup data.

  – **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

  NOTE

  Only MRS 3.1.0 or later supports data backup to OBS.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files.

The format of the backup file name is *Version_Data source_Task execution time*.**tar.gz**.

----End

9.11.2.2 Backing Up ClickHouse Metadata

Scenario

To ensure ClickHouse metadata security or before a major operation on ZooKeeper (such as upgrade or migration), you need to back up ClickHouse metadata. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up ClickHouse metadata. Both automatic and manual backup tasks are supported.

NOTICE

This function is supported only by MRS 3.1.0 or later.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path*/**LocalBackup/** has sufficient space on the active and standby management nodes.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the cluster to be operated from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task. **Periodic** indicates that the backup task is periodically executed. **Manual** indicates that the backup task is manually executed.

To create a periodic backup task, set the following parameters:

- **Started**: indicates the time when the task is started for the first time.
- **Period**: indicates the task execution interval. The options include **Hours** and **Days**.
- **Backup Policy**: Only **Full backup every time** is supported.

Step 6 In **Configuration**, select **ClickHouse** under **Metadata and other data**.

Step 7 Set **Path Type** of **ClickHouse** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files.

  The default storage directory is *Data storage path*/**LocalBackup/**, for example, **/srv/BigData/LocalBackup**.

  If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  This value option is available only after you configure the environment by referring to **How Do I Configure the Environment When I Create a ClickHouse Backup Task on FusionInsight Manager and Set the Path Type to RemoteHDFS?**.

  You also need to configure the following parameters:

  - **Destination NameService Name**: indicates the NameService name of the standby cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files. The format of the backup file name is *Data source_Task execution time***.tar.gz**.

----End

9.11.2.3 Backing Up ClickHouse Service Data

Scenario

To ensure ClickHouse service data security routinely or before a major operation on ClickHouse (such as upgrade or migration), you need to back up ClickHouse service data. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up ClickHouse service data. Both automatic and manual backup tasks are supported.

NOTICE

This function is supported only by MRS 3.1.0 or later.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- You have planned the backup type, period, object, and directory based on service requirements.
- The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the cluster to be operated from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 9-74 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	● **Full backup at the first time and incremental backup subsequently** ● **Full backup every time** ● **Full backup once every n times** NOTE ● Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported.

Step 6 In **Configuration**, select **ClickHouse** under **Service Data**.

Step 7 Set **Path Type** of **ClickHouse** to a backup directory type.

Currently, only the **RemoteHDFS** type is available.

RemoteHDFS: indicates that backup files are stored in HDFS of the standby cluster.

This value option is available only after you configure the environment by referring to **How Do I Configure the Environment When I Create a ClickHouse Backup Task on FusionInsight Manager and Set the Path Type to RemoteHDFS?**.

You also need to configure the following parameters:

- **Destination NameService Name**: indicates the NameService name of the standby cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
- **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
- **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
- **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a

snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.

- **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
- **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

Step 8 Set **Maximum Number of Recovery Points** to the number of snapshots that can be retained in the cluster.

Step 9 Set **Backup Content** to one or multiple ClickHouse tables to be backed up.

You can select backup data using either of the following methods:

- Adding a backup data file

  Click the name of a database in the navigation tree to show all the tables in the database, and select specified tables.

- Regular expression filtering

  a. Click **Query Regular Expression**.
  b. Enter the database where the ClickHouse tables are located in the first text box as prompted. The database must be the same as the existing database, for example, **default**.
  c. Enter a regular expression in the second text box. Standard regular expressions are supported. For example, to get all tables in the database, enter **([\s\S]*?)**. To get tables named in the format of letters and digits, for example, **tb1**, enter **tb\d***.
  d. Click **Refresh** to view the displayed tables in **Directory Name**.
  e. Click **Synchronize** to save the result.

NOTE

- When entering regular expressions, click + or − to add or delete an expression.
- If the selected table or directory is incorrect, click **Clear Selected Node** to deselect it.

Step 10 Click **Verify** to check whether the backup task is configured correctly.

The possible causes of the verification failure are as follows:

- The target NameNode IP address is incorrect.
- The directory or table to be backed up does not exist.
- The name of the NameService is incorrect.

Step 11 Click **OK**.

Step 12 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Data source_Task creation time*, and the subdirectory is used to save latest data source backup files.

----End

9.11.2.4 Backing Up DBService Data

Scenario

To ensure DBService service data security routinely or before a major operation on DBService (such as upgrade or migration), you need to back up DBService data. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up DBService data. Both automatic and manual backup tasks are supported.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path*/**LocalBackup/** has sufficient space on the active and standby management nodes.
- If you want to back up data to NAS, you have deployed the NAS server in advance.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the cluster to be operated from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 9-75 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	● **Full backup at the first time and incremental backup subsequently** ● **Full backup every time** ● **Full backup once every n times** **NOTE** ● Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. ● If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, select **DBService**.

NOTE

If there are multiple DBService services, all DBService services are backed up by default. You can click **Assign Service** to specify the services to be backed up.

Step 7 Set **Path Type** of **DBService** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files.

  The default storage directory is *Data storage path*/**LocalBackup/**, for example, **/srv/BigData/LocalBackup**.

  If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select this option, set the following parameters:

  – **Target Path**: indicates the HDFS directory for storing the backup files. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.

  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  - **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the source cluster.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.

  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Server Shared Path**: indicates the backup path on the SFTP server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **OBS**: indicates that backup files are stored in OBS.

  If you select this option, set the following parameters:
  - **Target Path**: indicates the OBS directory for storing backup data.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

    NOTE

    Only MRS 3.1.0 or later supports data backup to OBS.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files.

The format of the backup file name is *Version_Data source_Task execution time*.**tar.gz**.

----End

9.11.2.5 Backing Up HBase Metadata

Scenario

To ensure HBase metadata security (including tableinfo files and HFiles) or before a major operation on HBase system tables (such as upgrade or migration), you need to back up HBase metadata to prevent HBase service unavailability caused by HBase system table directory or file damages. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up HBase metadata. Both automatic and manual backup tasks are supported.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path***/LocalBackup/** has sufficient space on the active and standby management nodes.
- If you want to back up data to NAS, you have deployed the NAS server in advance.
- The **fs.defaultFS** parameter settings of HBase are the same as those of Yarn and HDFS.
- If HBase data is stored in the local HDFS, HBase metadata can be backed up to OBS. If HBase data is stored in OBS, data backup is not supported.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the cluster to be operated from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 9-76 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	● **Full backup at the first time and incremental backup subsequently** ● **Full backup every time** ● **Full backup once every n times** NOTE ● Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. ● If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, select **HBase** under **Metadata and other data**.

NOTE

If there are multiple HBase services, all HBase services are backed up by default. You can click **Assign Service** to specify the services to be backed up.

Step 7 Set **Path Type** of **HBase** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files.

  The default storage directory is *Data storage path*/**LocalBackup/**, for example, **/srv/BigData/LocalBackup**.

  If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  – **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the source cluster.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:

  – **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.

  – **Server IP Address**: indicates the IP address of the server where the backup data is stored.

  – **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.

  – **Username**: indicates the username for connecting to the server using the SFTP protocol.

  – **Password**: indicates the password for connecting to the server using the SFTP protocol.

  – **Server Shared Path**: indicates the backup path on the SFTP server.

  – **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **OBS**: indicates that backup files are stored in OBS.

  If you select this option, set the following parameters:

  – **Target Path**: indicates the OBS directory for storing backup data.

  – **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

  NOTE

  Only MRS 3.1.0 or later supports data backup to OBS.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files. The format of the backup file name is *Version_Data source_Task execution time*.**tar.gz**.

----End

9.11.2.6 Backing Up HBase Service Data

Scenario

To ensure HBase service data security routinely or before a major operation on HBase (such as upgrade or migration), you need to back up HBase service data. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up HBase service data. Both automatic and manual backup tasks are supported.

The following situations may occur during the HBase service data backup:

- When a user creates an HBase table, **KEEP_DELETED_CELLS** is set to **false** by default. When the user backs up this HBase table, deleted data will be backed up and junk data may exist after data restoration. This parameter can be set to **true** manually when an HBase table is created based on service requirements.
- When a user manually specifies the timestamp when writing data into an HBase table and the specified time is earlier than the last backup time of the HBase table, new data may not be backed up in incremental backup tasks.
- The HBase backup function cannot back up the access control lists (ACLs) for reading, writing, executing, creating, and managing HBase global or namespaces. After HBase data is restored, you need to reset the role permissions on FusionInsight Manager.
- If the backup data of the standby cluster is lost in an existing HBase backup task, the next incremental backup will fail, and you need to create an HBase backup task again. However, the next full backup task will be normal.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- Backup policies, including the backup task type, period, backup object, backup directory, and Yarn queue required by the backup task are planned based on service requirements.
- The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.
- On the HDFS client, you have executed the **hdfs lsSnapshottableDir** command as user **hdfs** to check the list of directories for which HDFS snapshots have been created in the current cluster and ensured that the HDFS parent directory or subdirectory where data files to be backed up are stored does not have HDFS snapshots. Otherwise, the backup task cannot be created.
- If you want to back up data to NAS, you have deployed the NAS server in advance.
- The **fs.defaultFS** parameter settings of HBase are the same as those of Yarn and HDFS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the cluster to be operated from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 9-77 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	● **Full backup at the first time and incremental backup subsequently** ● **Full backup every time** ● **Full backup once every n times** **NOTE** ● Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. ● If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, choose **HBase** > **HBase** under **Service data**.

Step 7 Set **Path Type** of **HBase** to a backup directory type.

The following backup directory types are supported:

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  – **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.

  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Server Shared Path**: indicates the backup path on the SFTP server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

Step 8 Set **Maximum Number of Recovery Points** to the number of snapshots that can be retained in the cluster.

Step 9 Set **Backup Content** to one or multiple HBase tables to be backed up.

You can select backup data using either of the following methods:

- Adding a backup data file

  Click the name of a database in the navigation tree to show all the tables in the database, and select specified tables.

- Selecting using regular expressions

  a. Click **Query Regular Expression**.
  b. Enter the namespace where the HBase tables are located in the first text box as prompted. The namespace must be the same as the existing namespace, for example, **default**.
  c. Enter a regular expression in the second text box. Standard regular expressions are supported. For example, to get all tables in the namespace, enter **([\s\S]*?)**. To get tables whose names consist of letters and digits, for example, **tb1**, enter **tb\d***.
  d. Click **Refresh** to view the displayed tables in **Directory Name**.
  e. Click **Synchronize** to save the result.

  NOTE

  – When entering regular expressions, click + or − to add or delete an expression.

  – If the selected table or directory is incorrect, click **Clear Selected Node** to deselect it.

Step 10 Click **Verify** to check whether the backup task is configured correctly.

The possible causes of the verification failure are as follows:

- The target NameNode IP address is incorrect.
- The queue name is incorrect.
- The parent directory or subdirectory of the HDFS directory where HBase table data files to be backed up are stored has HDFS snapshots.
- The directory or table to be backed up does not exist.

Step 11 Click **OK**.

Step 12 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Data source_Task creation time*, and the subdirectory is used to save latest data source backup files. All the backup file sets are stored in the related snapshot directories.

----End

9.11.2.7 Backing Up NameNode Data

Scenario

To ensure NameNode service data security routinely or before a major operation on NameNode (such as upgrade or migration), you need to back up NameNode data. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up NameNode data. Both automatic and manual backup tasks are supported.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path*/**LocalBackup/** has sufficient space on the active and standby management nodes.
- If you want to back up data to NAS, you have deployed the NAS server in advance.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the cluster to be operated from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 9-78 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.

Parameter	Description
Backup Policy	Only **Full backup every time** is supported. **NOTE** ● Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. ● If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, select **NameNode**.

Step 7 Set **Path Type** of **NameNode** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files. By default, the backup files are stored in *Data storage path*/**LocalBackup/**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster. If you select this option, set the following parameters:
  - **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the service plane IP address of the NameNode in the standby cluster.
  - **Target Path**: indicates the path for storing backup files.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol. If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol. If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.

  - **Server Shared Path**: indicates the backup path on the SFTP server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **OBS**: indicates that backup files are stored in OBS.

  If you select this option, set the following parameters:
  - **Target Path**: indicates the OBS directory for storing backup data.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

  NOTE

  Only MRS 3.1.0 or later supports data backup to OBS.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files.

The format of the backup file name is *Version_Data source_Task execution time***.tar.gz**.

----End

9.11.2.8 Backing Up HDFS Service Data

Scenario

To ensure HDFS service data security routinely or before a major operation on HDFS (such as upgrade or migration), you need to back up HDFS service data. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up HDFS service data. Both automatic and manual backup tasks are supported.

NOTE

Encrypted directories cannot be backed up or restored.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.

- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- Backup policies, including the backup task type, period, backup object, backup directory, and Yarn queue required by the backup task are planned based on service requirements.
- The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.
- On the HDFS client, you have executed the **hdfs lsSnapshottableDir** command as user **hdfs** to check the list of directories for which HDFS snapshots have been created in the current cluster and ensured that the HDFS parent directory or subdirectory where data files to be backed up are stored does not have HDFS snapshots. Otherwise, the backup task cannot be created.
- If you want to back up data to NAS, you have deployed the NAS server in advance.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the cluster to be operated from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 9-79 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.

Parameter	Description
Backup Policy	● **Full backup at the first time and incremental backup subsequently** ● **Full backup every time** ● **Full backup once every n times** **NOTE** ● Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. ● If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, select **HDFS**.

Step 7 Set **Path Type** of **HDFS** to a backup directory type.

The following backup directory types are supported:

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  – **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.

  – **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.

  – **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.

  – **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.

  – **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

  – **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.

  – **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.

  – **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol. If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.

  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Server Shared Path**: indicates the backup path on the SFTP server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

Step 8 Set **Maximum Number of Recovery Points** to the number of snapshots that can be retained in the cluster.

Step 9 Set **Backup Content** to one or multiple HDFS directories to be backed up based on service requirements.

You can select backup data using either of the following methods:

- Adding a backup data file

  Click the name of a database in the navigation tree to show all the tables in the database, and select specified tables.

- Selecting using regular expressions
  1. Click **Query Regular Expression**.
  2. Enter the parent directory full path of the directory in the first text box as prompted. The directory must be the same as the existing directory, for example, **/tmp**.
  3. Enter a regular expression in the second text box. Standard regular expressions are supported. For example, to get all files or subdirectories

in the parent directory, enter **([\s\S]*?)**. To get files whose names consist of letters and digits, for example, **file*1***, enter **file\d***.

d. Click **Refresh** to view the displayed directories in **Directory Name**.

e. Click **Synchronize** to save the result.

NOTE

- When entering regular expressions, click + or − to add or delete an expression.
- If the selected table or directory is incorrect, click **Clear Selected Node** to deselect it.
- The backup directory cannot contain files that have been written for a long time. Otherwise, the backup task will fail. Therefore, you are not advised to perform operations on the top-level directory, such as **/user**, **/tmp**, and **/mr-history**.

Step 10 Click **Verify** to check whether the backup task is configured correctly.

The possible causes of the verification failure are as follows:

- The target NameNode IP address is incorrect.
- The queue name is incorrect.
- The parent directory or subdirectory of the HDFS directory where data files to be backed up are stored has HDFS snapshots.
- The directory or table to be backed up does not exist.
- The name of the NameService is incorrect.

Step 11 Click **OK**.

Step 12 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Data source_Task creation time*, and the subdirectory is used to save latest data source backup files. All the backup file sets are stored in the related snapshot directories.

----End

9.11.2.9 Backing Up Hive Service Data

Scenario

To ensure Hive service data security routinely or before a major operation on Hive (such as upgrade or migration), you need to back up Hive service data. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up Hive service data. Both automatic and manual backup tasks are supported.

- Hive backup and restoration cannot identify the service and structure relationships of objects such as Hive tables, indexes, and views. When executing backup and restoration tasks, you need to manage unified restoration points based on service scenarios to ensure proper service running.

- Hive backup and restoration do not support Hive on RDB data tables. You need to back up and restore original data tables in external databases independently.
- If the backup data of the standby cluster is lost in an existing Hive backup task that contains Hive on HBase tables, the next incremental backup will fail, and you need to create a Hive backup task again. However, the next full backup task will be normal.
- After the backup function of FusionInsight Manager is used to back up the HDFS directories at the Hive table level, the Hive tables cannot be deleted and recreated.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- Backup policies, including the backup task type, period, backup object, backup directory, and Yarn queue required by the backup task are planned based on service requirements.
- The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.
- On the HDFS client, you have executed the **hdfs lsSnapshottableDir** command as user **hdfs** to check the list of directories for which HDFS snapshots have been created in the current cluster and ensured that the HDFS parent directory or subdirectory where data files to be backed up are stored does not have HDFS snapshots. Otherwise, the backup task cannot be created.
- If you want to back up data to NAS, you have deployed the NAS server in advance.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the cluster to be operated from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 9-80 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	● **Full backup at the first time and incremental backup subsequently** ● **Full backup every time** ● **Full backup once every n times** **NOTE** ● Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. ● If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, choose **Hive** > **Hive**.

Step 7 Set **Path Type** of **Hive** to a backup directory type.

The following backup directory types are supported:

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster. If you select this option, set the following parameters:
  - **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol. If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol. If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory,

and the user group and owner group of the shared path must be **nobody:nobody**.)

- **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
- **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
- **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Server Shared Path**: indicates the backup path on the SFTP server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

Step 8 Set **Maximum Number of Recovery Points** to the number of snapshots that can be retained in the cluster.

Step 9 Set **Backup Content** to one or multiple Hive tables to be backed up.

You can select backup data using either of the following methods:

- Adding a backup data file

  Click the name of a database in the navigation tree to show all the tables in the database, and select specified tables.

- Selecting using regular expressions
    a. Click **Query Regular Expression**.
    b. Enter the database where the Hive tables are located in the first text box as prompted. The database must be the same as the existing database, for example, **default**.
    c. Enter a regular expression in the second text box. Standard regular expressions are supported. For example, to get all tables in the database, enter **([\s\S]*?)**. To get tables whose names consist of letters and digits, for example, **tb1**, enter **tb\d***.
    d. Click **Refresh** to view the displayed tables in **Directory Name**.
    e. Click **Synchronize** to save the result.

    NOTE

    - When entering regular expressions, click + or − to add or delete an expression.
    - If the selected table or directory is incorrect, click **Clear Selected Node** to deselect it.

Step 10 Click **Verify** to check whether the backup task is configured correctly.

The possible causes of the verification failure are as follows:

- The target NameNode IP address is incorrect.
- The queue name is incorrect.
- The parent directory or subdirectory of the HDFS directory where data files to be backed up are stored has HDFS snapshots.
- The directory or table to be backed up does not exist.
- The name of the NameService is incorrect.

Step 11 Click **OK**.

Step 12 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Data source_Task creation time*, and the subdirectory is used to save latest data source backup files. All the backup file sets are stored in the related snapshot directories.

----End

9.11.2.10 Backing Up Kafka Metadata

Scenario

To ensure Kafka metadata security or before a major operation on ZooKeeper (such as upgrade or migration), you need to back up Kafka metadata. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up Kafka metadata. Both automatic and manual backup tasks are supported.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path***/LocalBackup/** has sufficient space on the active and standby management nodes.
- If you want to back up data to NAS, you have deployed the NAS server in advance.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the cluster to be operated from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 9-81 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.

Parameter	Description
Backup Policy	● **Full backup at the first time and incremental backup subsequently** ● **Full backup every time** ● **Full backup once every n times** **NOTE** ● Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. ● If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, select **Kafka**.

NOTE

If there are multiple Kafka services, all Kafka services are backed up by default. You can click **Assign Service** to specify the services to be backed up.

Step 7 Set **Path Type** of **Kafka** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files. By default, the backup files are stored in *Data storage path*/**LocalBackup/**.

  If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select this option, set the following parameters:

  – **Target Path**: indicates the HDFS directory for storing the backup files. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.

  – **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

  – **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  – **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **OBS**: indicates that backup files are stored in OBS.

  If you select this option, set the following parameters:

  – **Target Path**: indicates the OBS directory for storing backup data.

  – **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

  NOTE

  Only MRS 3.1.0 or later supports data backup to OBS.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files. The format of the backup file name is *Version_Data source_Task execution time***.tar.gz**.

----End

9.11.3 Recovering Data

9.11.3.1 Restoring Manager Data

Scenario

Manager data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in FusionInsight Manager, an exception occurs or the operation has not achieved the expected result. All modules are faulty and become unavailable.

System administrators can create a restoration task in FusionInsight Manager to recover Manager data. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that of data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the Manager data that is generated after the data backup and before the data restoration will be lost.

Impact on the System

- In the restoration process, the Controller needs to be restarted and FusionInsight Manager cannot be logged in or operated during the restart.

- In the restoration process, all clusters need to be restarted and cannot be accessed during the restart.
- After data restoration, the data, such as system configuration, user information, alarm information, and audit information, that is generated after the data backup and before the data restoration will be lost. This may result in data query failure or cluster access failure.
- After the Manager data is recovered, the system forces the LdapServer of each cluster to synchronize data from the OLadp.

Prerequisites

- To restore data from a remote HDFS, you need to prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, system mutual trust needs to be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The status of the OMS resources and the LdapServer instances of each cluster is normal. If the status is abnormal, data restoration cannot be performed.
- The status of the cluster hosts and services is normal. If the status is abnormal, data restoration cannot be performed.
- The cluster host topologies during data restoration and data backup are the same. If the topologies are different, data restoration cannot be performed and you need to back up data again.
- The services added to the cluster during data restoration and data backup are the same. If the topologies are different, data restoration cannot be performed and you need to back up data again.
- The upper-layer applications that depend on the cluster are stopped.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view the historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restore** > **Restoring Management**. On the displayed page, click **Create**.

Step 4 Set **Task Name** to the name of the restoration task.

Step 5 Set **Recovery Object** to **OMS**.

Step 6 Select **OMS**.

Step 7 Set **Path Type** of **OMS** to a backup directory type.

The settings vary according to backup directory types:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node.

  If you select **LocalDir**, you also need to set **Source Path** to select the backup file to be restored, for example, *Version_Data source_Task execution time*.**tar.gz**.
- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select **LocalHDFS**, set the following parameters:
  - **Source Path**: indicates the full path of the backup file in the HDFS, for example, *Backup path/Backup task name_Task creation time/ Version_Data source_Task execution time*.**tar.gz**.
  - **Cluster for Restoration**: Enter the name of the cluster used during restoration task execution.
  - **Source NameService Name**: indicates the NameService name that corresponds to the backup directory when a restoration task is executed. The default value is **hacluster**.
- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select **RemoteHDFS**, set the following parameters:
  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time*.**tar.gz**.
  - **Source Cluster**: Select the cluster of the Yarn queue used by the recovery data.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol. If you select **NFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the complete path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol. If you select **CIFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select **SFTP**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **OBS**: indicates that backup files are stored in OBS.

  If you select **OBS**, set the following parameters:
  - **Source Path**: indicates the full OBS path of a backup file, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

    NOTE

    Only MRS 3.1.0 or later supports saving backup files in OBS.

Step 8 Click **OK**.

Step 9 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

Step 10 Log in to the active and standby management nodes as user **omm** using PuTTY.

Step 11 Run the following command to restart OMS:

sh ${BIGDATA_HOME}/om-server/om/sbin/restart-oms.sh

The command is run successfully if the following information is displayed:

```
start HA successfully.
```

Run **sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh** to check whether **HAAllResOK** of the management node is **Normal** and whether FusionInsight Manager can be logged in again. If yes, OMS is restarted successfully.

Step 12 On FusionInsight Manager, click **Cluster**, click the name of the target cluster, and choose **Services** > **KrbServer**. On the displayed page, choose **More** > **Synchronize Configuration**, click **OK**, and wait for the KrbServer configuration to be synchronized and the service to be restarted.

Step 13 Choose **Cluster**, click the name of the desired cluster, and choose **More** > **Synchronize Configurations**, click **OK**, and wait until the cluster configuration is synchronized successfully.

Step 14 On FusionInsight Manager, click **Cluster**, click the name of the target cluster, and choose **More** > **Restart**. On the displayed page, enter the password of the current login user, click **OK**, and wait for the cluster to be restarted.

----End

9.11.3.2 Restoring ClickHouse Metadata

Scenario

ClickHouse metadata needs to be restored in the following scenarios: Data is modified or deleted unexpectedly and needs to be restored. After a user performs major operations (such as upgrade and migration) on ZooKeeper, an exception occurs or the expected result is not achieved. The ClickHouse component is faulty and becomes unavailable. Data is migrated to a new cluster.

Users can create a ClickHouse restoration task on FusionInsight Manager. Only manual restoration tasks are supported.

NOTICE

- This function is supported only by MRS 3.1.0 or later.
- Data restoration can be performed only when the system version is consistent with that during data backup.
- To restore ClickHouse metadata when the service is running properly, you are advised to manually back up the latest ClickHouse metadata before restoration. Otherwise, the ClickHouse metadata that is generated after the data backup and before the data restoration will be lost.
- ClickHouse metadata restoration and service data restoration cannot be performed at the same time. Otherwise, service data restoration fails. You are advised to restore service data after metadata restoration is complete.

Impact on the System

- After the metadata is restored, the data generated after the data backup and before the data restoration is lost.
- After the metadata is restored, the ClickHouse upper-layer applications need to be started.

Prerequisites

- You have checked the path for storing ClickHouse metadata backup files.
- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of the specified task in the task list, choose **More** > **View History**.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.
- **Backup Path**: indicates the full path where backup files are stored.

  Select the correct path, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the cluster to be operated from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **ClickHouse** under **Metadata and other data**.

Step 8 Set **Path Type** of **ClickHouse** to a restoration directory type.

The configurations vary based on backup directory types:

- **LocalDir**: indicates that data is restored from the local disk of the active management node.

  If you select this value, you also need to configure the following parameters:

  - **Source Path**: backup file to be restored, for example, *Backup task name_Data source_Task execution time***.tar.gz**.
  - **Logical Cluster**: Enter the ClickHouse logical cluster whose data has been backed up.

- **RemoteHDFS**: indicates that data is restored from the HDFS directory of the standby cluster.

  If you select this value option, you also need to configure the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Data source_Task execution time***.tar.gz**.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate the row where the created task is located, and click **Start** in the **Operation** column. In the displayed dialog box, click **OK** to start the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

Step 11 Choose **Cluster** > **Services** and start the ClickHouse service.

----End

9.11.3.3 Restoring ClickHouse Service Data

Scenario

ClickHouse data needs to be restored in the following scenarios: Data is modified or deleted unexpectedly and needs to be restored. After a user performs major operations (such as upgrade and migration) on ClickHouse, an exception occurs or the expected result is not achieved. All modules are faulty and become unavailable. Data is migrated to a new cluster.

Users can create a ClickHouse restoration task on FusionInsight Manager to restore data. Only manual restoration tasks are supported.

The ClickHouse backup and restoration functions cannot identify the service and structure relationships of objects such as ClickHouse tables, indexes, and views. When executing backup and restoration tasks, you need to manage unified restoration points based on service scenarios to ensure proper service running.

NOTICE

- This function is supported only by MRS 3.1.0 or later.
- Data restoration can be performed only when the system version is consistent with that during data backup.
- To restore the data when services are normal, manually back up the latest management data first and then restore the data. Otherwise, the ClickHouse data that is generated after the data backup and before the data restoration will be lost.
- ClickHouse metadata restoration and service data restoration cannot be performed at the same time. Otherwise, service data restoration fails. You are advised to restore service data after metadata restoration is complete.

Impact on the System

- During data restoration, user authentication stops and users cannot create new connections.
- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is restored, the ClickHouse upper-layer applications need to be started.

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.

- The database for storing restored data tables, the HDFS save path of data tables, and the list of users who can access restored data are planned.
- The ClickHouse backup file save path is correct.
- The ClickHouse upper-layer applications are stopped.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the row where the specified backup task is located, choose **More** > **View History** in the **Operation** column to display the historical execution records of the backup task.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.
- **Backup Path**: indicates the full path where backup files are stored.

  Select the correct path, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the cluster to be operated from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **ClickHouse** under **Service data**.

Step 8 Set **Path Type** of **ClickHouse** to a backup directory type.

Currently, the backup directory supports only the **RemoteHDFS** type.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster. If you select this value option, you also need to configure the following parameters:
  - **Source NameService Name**: indicates the NameService name of the backup data cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - Source Path: indicates the full path of HDFS directory for storing backup data of the standby cluster. For details, see the **Backup Path** obtained in

step 2, for example, *Backup path/Backup task name_Data source_Task creation time*.

- **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
- **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate the row where the created task is located, and click **Start** in the **Operation** column.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

----End

9.11.3.4 Restoring DBService data

Scenario

DBService data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in DBService, an exception occurs or the operation has not achieved the expected result. All modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a recovery task in FusionInsight Manager to recover DBService data. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that of data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the DBService data that is generated after the data backup and before the data recovery will be lost.
- By default, MRS clusters use DBServices to store metadata of Hive, Hue, Loader, Spark, and Oozie. Restoring DBService data will restore the metadata of all these components.

Impact on the System

- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is restored, the configurations of the components that depend on DBService may expire and these components need to be restarted.

Prerequisites

- To restore data from a remote HDFS, you need to prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The status of the active and standby DBService instances is normal. If the status is abnormal, data restoration cannot be performed.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view the historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the cluster to be operated from **Recovery Object**.

Step 7 In the **Restoration Configuration** area, select **DBService**.

NOTE

If multiple DBServices are installed, select the DBServices to be restored.

Step 8 Set **Path Type** of **DBService** to a backup directory type.

The settings vary according to backup directory types:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node.

  If you select **LocalDir**, you also need to set **Source Path** to select the backup file to be restored, for example, *Version_Data source_Task execution time*.**tar.gz**.

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select **LocalHDFS**, set the following parameters:

  - **Source Path**: indicates the full path of the backup file in the HDFS, for example, *Backup path/Backup task name_Task creation time/ Version_Data source_Task execution time***.tar.gz**.
  - **Source NameService Name**: indicates the NameService name that corresponds to the backup directory when a restoration task is executed. The default value is **hacluster**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select **RemoteHDFS**, set the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select **NFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select **CIFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.

  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select **SFTP**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **OBS**: indicates that backup files are stored in OBS.

  If you select **OBS**, set the following parameters:
  - **Source Path**: indicates the full OBS path of a backup file, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

    NOTE

    Only MRS 3.1.0 or later supports saving backup files in OBS.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

----End

9.11.3.5 Restoring HBase Metadata

Scenario

To ensure HBase metadata security (including tableinfo files and HFiles) or before a major operation on HBase system tables (such as upgrade or migration), you need to back up HBase metadata to prevent HBase service unavailability caused by HBase system table directory or file damages. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

System administrators can create a recovery task in FusionInsight Manager to recover HBase metadata. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the HBase data that is generated after the data backup and before the data recovery will be lost.
- It is recommended that a data restoration task restore the metadata of only one component to prevent the data restoration of other components from being affected by stopping a service or instance. If data of multiple components is restored at the same time, data restoration may fail.

  HBase metadata cannot be restored at the same time as NameNode metadata. As a result, data restoration fails.

Impact on the System

- Before restoring the metadata, you need to stop the HBase service, during which the HBase upper-layer applications are unavailable.
- After the metadata is restored, the data generated after the data backup and before the data restoration is lost.
- After the metadata is restored, the upper-layer applications of HBase need to be started.

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- You have checked the path for storing HBase metadata backup files.

- The HBase service has been stopped before its metadata is restored.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view the historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the cluster to be operated from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **HBase** under **Metadata and other data**.

NOTE

If multiple HBase services are installed, select the HBase services to be restored.

Step 8 Set **Path Type** of **HBase** to a backup directory type.

The settings vary according to backup directory types:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node.

  If you select **LocalDir**, you also need to set **Source Path** to select the backup file to be restored, for example, *Version_Data source_Task execution time***.tar.gz**.
- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select **RemoteHDFS**, set the following parameters:
  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.

  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select **NFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

- **CIFS**: indicates that backup files are stored in NAS using the CIFS protocol.

  If you select **CIFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select **SFTP**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.

  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time*.**tar.gz**.
- **OBS**: indicates that backup files are stored in OBS.

  If you select **OBS**, set the following parameters:

  - **Source Path**: indicates the full OBS path of a backup file, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time*.**tar.gz**.

    NOTE

    Only MRS 3.1.0 or later supports saving backup files in OBS.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

----End

9.11.3.6 Restoring HBase Service Data

Scenario

HBase data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in HBase, an exception occurs or the operation has not achieved the expected result. All modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a recovery task in FusionInsight Manager to recover HBase data. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the HBase data that is generated after the data backup and before the data recovery will be lost.

Impact on the System

- During the data recovery process, the system disables the HBase table to be recovered and the table cannot be accessed in this moment. The data recovery process takes several minutes, during which the HBase upper-layer applications are unavailable.
- During data restoration, user authentication stops and users cannot create new connections.
- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is recovered, the HBase upper-layer applications need to be started.

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The directory for saving the backup file has been checked.
- The HBase upper-layer applications have been stopped.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the cluster to be operated from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **HBase** under **Service Data**.

Step 8 Set **Path Type** of **HBase** to a backup directory type.

The following backup directory types are supported:

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster. If you select **RemoteHDFS**, set the following parameters:
  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of the backup file in the HDFS, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol. If you select **NFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.

  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol. If you select **CIFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select **SFTP**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.

  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

Step 9 Set **Backup Data** column in **Data Configuration** to one or multiple backup data sources to be recovered. In the **Target Namespace** column, specify the target naming space after backup data recovery.

You are advised to set **Target Namespace** to a location that is different from the backup naming space.

Step 10 Set **Force recovery** to **true**, which indicates to forcibly recover all backup data when a data table with the same name already exists. If the data table contains new data added after backup, the new data will be lost after the data recovery. If you set the parameter to **false**, the restoration task is not executed if a data table with the same name exists.

Step 11 Click **Verify** to check whether the restoration task is configured correctly.

- If the queue name is incorrect, the verification fails.
- If the specified naming space does not exist, the verification fails.
- If the forcibly replacement conditions are not met, the verification fails.

Step 12 Click **OK** to save the settings.

Step 13 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

Step 14 Check whether HBase data is restored in an environment where HBase is newly installed or reinstalled.

- If yes, the administrator needs to set new permission for roles on FusionInsight Manager based on the original service plan.
- If no, no further operation is required.

----End

9.11.3.7 Restoring NameNode Data

Scenario

NameNode data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in NameNode, an exception occurs or the operation has not achieved the expected result. All modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a recovery task in FusionInsight Manager to recover NameNode data. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the NameNode data that is generated after the data backup and before the data recovery will be lost.
- It is recommended that a data restoration task restore the metadata of only one component to prevent the data restoration of other components from being affected by stopping a service or instance. If data of multiple components is restored at the same time, data restoration may fail.

  HBase metadata cannot be restored at the same time as NameNode metadata. As a result, data restoration fails.

Impact on the System

- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is recovered, the NameNode needs to be restarted and is unavailable during the restart.
- After data is restored, metadata and service data may not be matched, the HDFS enters the security mode, and the HDFS service fails to be started. .

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.
- On FusionInsight Manager, all the NameNode role instances whose data is to be recovered are stopped. Other HDFS role instances must keep running. After data is recovered, the NameNode role instances need to be restarted. The NameNode role instances cannot be accessed during the restart.
- The NameNode backup files are stored ***Data path*/LocalBackup/** on the active management node.

Procedure

Step 1 On FusionInsight Manager, click **Cluster**, click the name of the desired cluster, and choose **Services** > **HDFS**. On the displayed page, click **Instance** and click

NameNode to check whether the NameNode instances of the data to be restored are stopped. If the NameNode instances are not stopped, stop them.

Step 2 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 3 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 4 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 5 Click **Create**.

Step 6 Set **Task Name** to the name of the restoration task.

Step 7 Select the cluster to be operated from **Recovery Object**.

Step 8 In the **Restoration Configuration** area, select **NameNode**.

Step 9 Set **Path Type** of **NameNode** to a backup directory type.

The settings vary according to backup directory types:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node.

  If you select **LocalDir**, set the following parameters:

  – **Source Path**: indicates the full path of the backup file on the local disk, for example, *Backup path/Backup task name_Task creation time/ Version_Data source_Task execution time***.tar.gz**.

  – **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select **RemoteHDFS**, set the following parameters:

  – **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.

  – **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.

  – **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.

  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol. If you select **NFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol. If you select **CIFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select **SFTP**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.

  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **OBS**: indicates that backup files are stored in OBS.

  If you select **OBS**, set the following parameters:
  - **Source Path**: indicates the full OBS path of a backup file, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

    NOTE

    Only MRS 3.1.0 or later supports saving backup files in OBS.

Step 10 Click **OK**.

Step 11 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

Step 12 On FusionInsight Manager, click **Cluster**, click the name of the desired cluster, and choose **Services** > **HDFS**. On the displayed page, click **Configurations** and click **All Configurations**.

On the displayed page, enter the password of the administrator who has logged in for authentication and click **OK**. After the system displays "Operation succeeded", click **Finish**. The service is started successfully.

----End

9.11.3.8 Restoring HDFS Service Data

Scenario

HDFS data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in the HDFS, an exception occurs or the operation has not achieved the expected result. All modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a recovery task in FusionInsight Manager to recover HDFS data. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the HDFS data that is generated after the data backup and before the data recovery will be lost.
- The HDFS restoration operation cannot be performed for the directories used by running Yarn tasks, for example, **/tmp/logs**, **/tmp/archived**, and **/tmp/hadoop-yarn/staging**. Otherwise, data restoration using Distcp tasks fails due to file loss.

Impact on the System

- During data restoration, user authentication stops and users cannot create new connections.
- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is recovered, the HDFS upper-layer applications need to be started.

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The HDFS backup file save path is correct.
- The HDFS upper-layer applications are stopped.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the cluster to be operated from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **HDFS** under **Service Data**.

Step 8 Set **Path Type** of **HDFS** to a backup directory type.

The following backup directory types are supported:

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select **RemoteHDFS**, set the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **NFS**: indicates that backup files are stored in NAS using the NFS protocol. If you select **NFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.

  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **CIFS**: indicates that backup files are stored in NAS using the CIFS protocol. If you select **CIFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select **SFTP**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.

- **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
- **Username**: indicates the username for connecting to the server using the SFTP protocol.
- **Password**: indicates the password for connecting to the server using the SFTP protocol.
- **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
- **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
- **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
- **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

Step 9 In the **Backup Data** column of the **Data Configuration** page, select one or more pieces of backup data that needs to be restored based on service requirements. In the **Target Path** column, specify the target location after backup data restoration.

You are advised to set **Target Path** to a new path that is different from the backup path.

Step 10 Click **Verify** to check whether the restoration task is configured correctly.

- If the queue name is incorrect, the verification fails.
- If the specified directory to be restored does not exist, the verification fails.

Step 11 Click **OK**.

Step 12 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

----End

9.11.3.9 Restoring Hive Service Data

Scenario

Hive data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in the Hive, an exception occurs or the operation has not achieved the expected result. All modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a recovery task in FusionInsight Manager to recover Hive data. Only manual restoration tasks are supported.

Hive backup and restoration cannot identify the service and structure relationships of objects such as Hive tables, indexes, and views. When executing backup and restoration tasks, you need to manage unified restoration points based on service scenarios to ensure proper service running.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the Hive data that is generated after the data backup and before the data recovery will be lost.

Impact on the System

- During data restoration, user authentication stops and users cannot create new connections.
- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is recovered, the Hive upper-layer applications need to be started.

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The database for storing restored data tables, the HDFS save path of data tables, and the list of users who can access restored data are planned.
- The Hive backup file save path is correct.
- The Hive upper-layer applications are stopped.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the cluster to be operated from **Recovery Object**.

Step 7 In the **Restoration Configuration** area, select **Hive**.

Step 8 Set **Path Type** of **Hive** to a backup directory type.

The following backup directory types are supported:

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster. If you select **RemoteHDFS**, set the following parameters:
  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select a Hive backup file set that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **NFS**: indicates that backup files are stored in NAS using the NFS protocol. If you select **NFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select a Hive backup file set that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **CIFS**: indicates that backup files are stored in NAS using the CIFS protocol. If you select **CIFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select a Hive backup file set that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.

  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select **SFTP**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **1**.

Step 9 Set **Backup Data** in the **Data Configuration** to one or multiple backup data sources to be recovered based on service requirements. In the **Target Database** and **Target Path** columns, specify the target database and file save path after backup data recovery.

Configuration restrictions:

- Data can be restored to the original database, but data tables must be stored in a new path that is different from the backup path.
- To restore Hive index tables, select the Hive data tables that correspond to the Hive index tables to be restored.
- If a new restoration directory is selected to avoid affecting the current data, HDFS permission must be manually granted so that users who have permission of backup tables can access this directory.
- Data can be restored to other databases. In this case, HDFS permission must be manually granted so that users who have permission of backup tables can access the HDFS directory that corresponds to the database.

Step 10 Set **Force recovery** to **true**, which indicates to forcibly recover all backup data when a data table with the same name already exists. If the data table contains new data added after backup, the new data will be lost after the data recovery. If you set the parameter to **false**, the restoration task is not executed if a data table with the same name exists.

Step 11 Click **Verify** to check whether the restoration task is configured correctly.

- If the queue name is incorrect, the verification fails.
- If the specified directory to be restored does not exist, the verification fails.
- If the forcibly replacement conditions are not met, the verification fails.

Step 12 Click **OK**.

Step 13 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

----End

9.11.3.10 Restoring Kafka Metadata

Scenario

Kafka data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in ZooKeeper, an exception occurs or the operation has not achieved the expected result. All Kafka modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a recovery task in FusionInsight Manager to recover Kafka data. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To restore Kafka metadata when the service is running properly, you are advised to manually back up the latest Kafka metadata before restoration. Otherwise, the Kafka metadata that is generated after the data backup and before the data restoration will be lost.

Impact on the System

- After the metadata is restored, the data generated after the data backup and before the data restoration is lost.
- After the metadata is restored, the offset information stored on ZooKeeper by Kafka consumers is rolled back, resulting in repeated consumption.

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The Kafka service is disabled first, and then enabled upon data restoration.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the cluster to be operated from **Recovery Object**.

Step 7 In the **Restoration Configuration** area, select **Kafka**.

NOTE

If multiple Kafka services are installed, select the Kafka services to be restored.

Step 8 Set **Path Type** of **Kafka** to a backup directory type.

The settings vary according to backup directory types:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node.

  If you select **LocalDir**, you also need to set **Source Path** to select the backup file to be restored, for example, *Version_Data source_Task execution time*.**tar.gz**.

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select **LocalHDFS**, set the following parameters:

  - **Source Path**: indicates the full path of the backup file in the HDFS, for example, *Backup path/Backup task name_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Source NameService Name**: indicates the NameService name that corresponds to the backup directory when a restoration task is executed. The default value is **hacluster**.
- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select **RemoteHDFS**, set the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
- **NFS**: indicates that backup files are stored in NAS using the NFS protocol.

  If you select **NFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **CIFS**: indicates that backup files are stored in NAS using the CIFS protocol.

  If you select **CIFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.

  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time*.**tar.gz**.

- **OBS**: indicates that backup files are stored in OBS.

  If you select **OBS**, set the following parameters:

  - **Source Path**: indicates the full OBS path of a backup file, for example, *Backup path/Backup task name_Data source_Task creation time/ Version_Data source_Task execution time*.**tar.gz**.

    NOTE

    Only MRS 3.1.0 or later supports saving backup files in OBS.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

NOTICE

- If the Kafka service is deleted after the backup is complete, reinstall the Kafka service, restore its metadata, and restart the Kafka service. It is found that the Broker service cannot be started. In this case, the **/var/log/Bigdata/kafka/broker/server.log** file contains an error. An error example is as follows:

```
ERROR Fatal error during KafkaServer startup. Prepare to shutdown
(kafka.server.KafkaServer)kafka.common.InconsistentClusterIdException: The Cluster ID
kVSgfurUQFGGpHMTBqBPiw doesn't match stored clusterId Some(0Qftv9yBTAmf2iDPSlIk7g) in
meta.properties. The broker is trying to join the wrong cluster. Configured zookeeper.connect may
be wrong. at kafka.server.KafkaServer.startup(KafkaServer.scala:220) at
kafka.server.KafkaServerStartable.startup(KafkaServerStartable.scala:44) at kafka.Kafka
$.main(Kafka.scala:84) at kafka.Kafka.main(Kafka.scala)
```

  Check the value of **log.dirs** in the Kafka Broker configuration file **${BIGDATA_HOME}/Fusionsight_Current/*Broker/etc/server.properties**. The value is the Kafka data directory. Go to the Kafka data directory and change the value **0Qftv9yBTAmf2iDPSlIk7g** of **cluster.id** in **meta.properties** to **kVSgfurUQFGGpHMTBqBPiw** (the latest value in the error log).

- The preceding modification must be performed on each node where Broker is located. After the modification, restart the Kafka service.

----End

9.11.4 Enabling Cross-Cluster Replication

Scenario

DistCp is used to replicate the data stored in HDFS from a cluster to another cluster. DistCp depends on the cross-cluster replication function, which is disabled by default. You need to enable it for both clusters.

This section describes how to modify parameters on FusionInsight Manager to enable the cross-cluster replication function. After this function is enabled, you can create a backup task for backing up data to the remote HDFS (RemoteHDFS).

Impact on the System

Yarn needs to be restarted to enable the cross-cluster replication function and cannot be accessed during restart.

Prerequisites

- The **hadoop.rpc.protection** parameter of HDFS in the two clusters for data replication must use the same data transmission mode. The default value is **privacy**, indicating encrypted transmission. The value **authentication** indicates that transmission is not encrypted.
- For clusters in security mode, you need to configure mutual trust between clusters.

Procedure

Step 1 Log in to FusionInsight Manager of one of the two clusters.

Step 2 Choose **Cluster** > ***Name of the desired cluster* > Services** > **Yarn** > **Configurations**, and click **All Configurations**.

Step 3 In the navigation pane, choose **Yarn** > **Distcp**.

Step 4 Modify **dfs.namenode.rpc-address**, set **haclusterX.remotenn1** to the service IP address and RPC port of one NameNode instance of the peer cluster, and set **haclusterX.remotenn2** to the service IP address and RPC port number of the other NameNode instance of the peer cluster.

haclusterX.remotenn1 and **haclusterX.remotenn2** do not distinguish active and standby NameNodes. The default NameNode RPC port is 8020 and cannot be modified on Manager.

Examples of modified parameter values: **10.1.1.1:8020** and **10.1.1.2:8020**.

NOTE

- If data of the current cluster needs to be backed up to the HDFS of multiple clusters, you can configure the corresponding NameNode RPC addresses to haclusterX1, haclusterX2, haclusterX3, and haclusterX4.

Step 5 Click **Save**. In the confirmation dialog box, click **OK**.

Step 6 Restart the Yarn service.

Step 7 Log in to FusionInsight Manager of the other cluster and repeat **Step 2** to **Step 6**.

----End

9.11.5 Managing Local Quick Restoration Tasks

Scenario

When DistCp is used to back up data, the backup snapshot is saved to HDFS of the active cluster. FusionInsight Manager supports using the local snapshot for quick data restoration, requiring less time than restoring data from the standby cluster.

Use FusionInsight Manager and the snapshots on HDFS of the active cluster to create a local quick restoration task and execute the task.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the backup task list, locate a created task and click **Restore** in the **Operation** column.

Step 3 Check whether the system displays "No data is available for quick restoration. Create a task on the restoration management page to restore data".

- If yes, click **OK** to close the dialog box. No backup data snapshot is created in the active cluster, and no further action is required.
- If no, go to **Step 4** to create a local quick restoration task.

NOTE

Metadata does not support quick restoration.

Step 4 Set **Name** to the name of the local quick restoration task.

Step 5 Set **Configuration** to a data source.

Step 6 Set **Recovery Point List** to a recovery point that contains the backup data.

Step 7 Set **Queue Name** to the name of the Yarn queue used in the task execution. The name must be the same as the name of the queue that is running properly in the cluster.

Step 8 Set **Data Configuration** to the object to be recovered.

Step 9 Click **Verify**, and wait for the system to display "The restoration task configuration is verified successfully."

Step 10 Click **OK**.

Step 11 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

After the task is complete, **Task Status** of the task is displayed as **Successful**.

----End

9.11.6 Modifying a Backup Task

Scenario

This section describes how to modify the parameters of a created backup task on FusionInsight Manager to meet changing service requirements. The parameters of restoration tasks can only be viewed but cannot be modified.

Impact on the System

After a backup task is modified, the new parameters take effect when the task is executed next time.

Prerequisites

- A backup task has been created.
- A new backup task policy has been planned based on the actual situation.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the task list, locate a specified task, click **Configure** in the **Operation** column to go to the configuration modification page.

On the displayed page, modify the following parameters:

- Started
- Period
- Destination NameService Name
- Target NameNode IP Address
- Target Path
- Max Number of Backup Copies
- Maximum Number of Recovery Points
- Maximum Number of Maps
- Maximum Bandwidth of a Map

NOTE

After the **Target Path** parameter of a backup task is modified, this task will be performed as a full backup task for the first time by default.

Step 3 Click **OK** to save the settings.

----End

9.11.7 Viewing Backup and Restoration Tasks

Scenario

This section describes how to view created backup and recovery tasks and check their running status on FusionInsight Manager.

Prerequisites

You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration**.

Step 2 Click **Backup Management** or **Restoration Management**.

Step 3 In the task list, obtain the previous execution result in the **Task Status** and **Task Progress** column. Green indicates that the task is executed successfully, and red indicates that the execution fails.

Step 4 In the **Operation** column of a specified task in the task list, choose **More** > **View History** or click **View History** to view the historical record of backup and restoration task execution.

In the displayed window, click ⌄ before a specified record to display log information about the execution.

----End

Related Tasks

- Starting a backup or restoration task

  In the task list, locate a specified task and choose **More** > **Back Up Now** or click **Start** in the **Operation** column to start a backup or restoration task that is ready or fails to be executed. Executed restoration tasks cannot be repeatedly executed.

- Stopping a backup or restoration task

  In the task list, locate a specified task and choose **More** > **Stop** or click **Stop** in the **Operation** column to stop a backup or restoration task that is running. After the task is successfully stopped, its **Task Status** changes to **Stopped**.

- Deleting a backup or restoration task

  In the task list, locate a specified task and choose **More** > **Delete** or click **Delete** in the **Operation** column to delete a backup or restoration task. Backup data will be reserved by default after a task is deleted.

- Suspending a backup task

  In the task list, locate a specified task and choose **More** > **Suspend** in the **Operation** column to suspend a backup task. Only periodic backup tasks can be suspended. Suspended backup tasks are no longer executed automatically. When you suspend a backup task that is being executed, the task execution stops. To resume a task, choose **More** > **Resume**.

9.11.8 How Do I Configure the Environment When I Create a ClickHouse Backup Task on FusionInsight Manager and Set the Path Type to RemoteHDFS?

NOTE

This section applies only to MRS 3.1.0.

Question

How do I configure the environment when I create a ClickHouse backup task on FusionInsight Manager and set the path type to RemoteHDFS?

Answer

Step 1 Log in to FusionInsight Manager of the standby cluster.

Step 2 Choose **Cluster** > **Services** > **HDFS** and choose **More** > **Download Client**. Set **Select Client Type** to **Configuration Files Only**, select **x86_64** for x86 or **aarch64** for ARM based on the type of the node where the client is to be installed, and click **OK**.

Step 3 After the client file package is generated, download the client to the local PC as prompted and decompress the package.

For example, if the client file package is **FusionInsight_Cluster_1_HDFS_Client.tar**, decompress it to obtain **FusionInsight_Cluster_1_HDFS_ClientConfig_ConfigFiles.tar**, and then decompress **FusionInsight_Cluster_1_HDFS_ClientConfig_ConfigFiles.tar** to the **D:\FusionInsight_Cluster_1_HDFS_ClientConfig_ConfigFiles** directory on the local PC. The directory name cannot contain spaces.

Step 4 Go to the **FusionInsight_Cluster_1_HDFS_ClientConfig_ConfigFiles** client directory and obtain the **hosts** file.

Step 5 Log in to FusionInsight Manager of the source cluster.

Step 6 Choose **Cluster** > **Services** > **ClickHouse**, click **Instance**, and view the instance IP address of **ClickHouseServer**.

Step 7 Log in to the host nodes of the ClickHouseServer instances as user **root** and check whether the **/etc/hosts** file contains the host information in **Step 4**. If not, add the host information in **Step 4** to the **/etc/hosts** file.

----End

9.12 Security Management

9.12.1 Security Overview

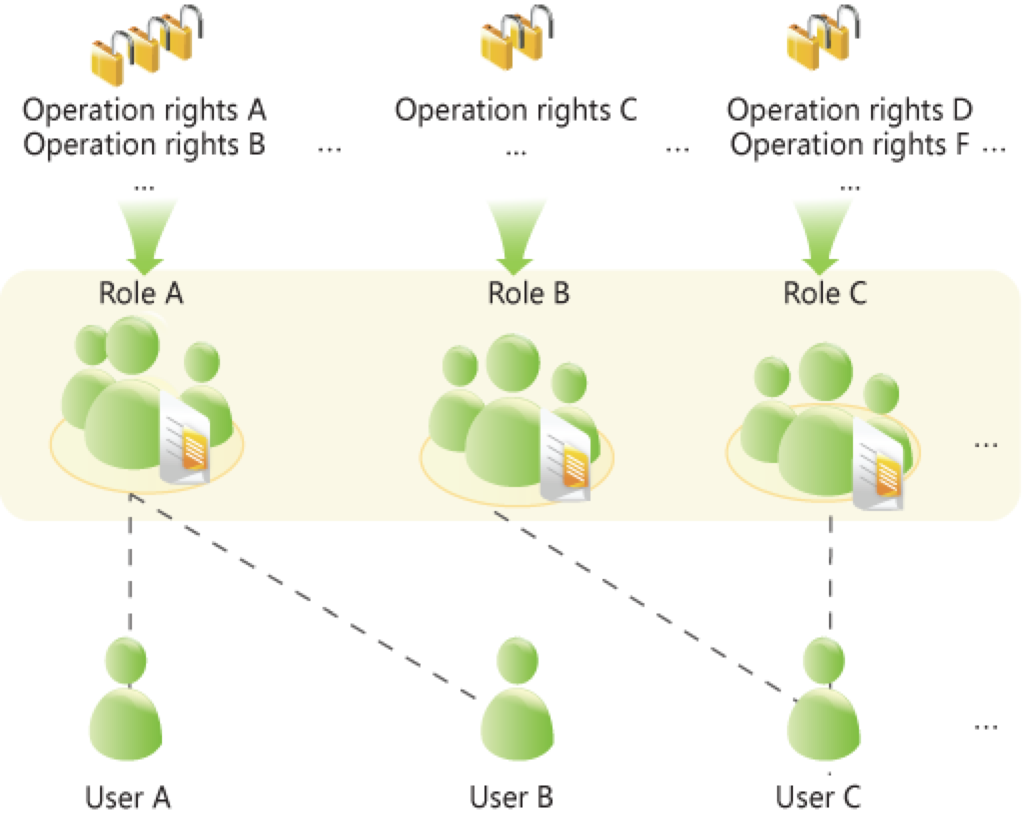
9.12.1.1 Right Model

Role-based Access Control

FusionInsight adopts the role-based access control (RBAC) mode to manage rights on the big data system. It integrates the right management functions of the components to centrally manage rights. Common users are shielded from internal right management details, and the right management operations are simplified for administrators, improving right management usability and user experience.

The right model of FusionInsight consists four parts, that is users, user groups, roles, and rights.

Figure 9-14 Right model



- **Right**

  Right, which is defined by components, allows users to access a certain resource of one component. Different components have different rights for their resources.

  For example:

  – HDFS provides read, write, and execute permissions on files.

  – HBase provides create, read, and write permissions on tables.

- **Role**

  Role is a collection of component rights. Each role can have multiple rights of multiple components. Different roles can have the rights of a resource of one component.

- **User group**

  User group is a collection of users. When a user group is bound to a role, users in this group obtain the rights defined by the role.

  Different user groups can be associated with the same role. A user group can also be associated with no role, and this user group does not have the rights of any component resources.

  NOTE

  In some components, the system grants related rights to specific user groups by default.

- **User**

  A user is a visitor to the system. Each user has the rights of the user group and role associated with the user. Users need to be added to the user group or associated with roles to obtain the corresponding rights.

Policy-based Access Control

The Ranger component uses policy-based access control (PBAC) to manage rights and implement fine-grained data access control on components such as HDFS, Hive, and HBase.

NOTE

The component supports only one right control mechanism. After the Ranger right control policy is enabled for the component, the right on the component in the role created on FusionInsight Manager becomes invalid (The ACL rules of HDFS and Yarn still take effect). You need to add a policy on the Ranger management page to grant rights on resources.

The Ranger right model consists of multiple right policies. A right policy consists of the following parts:

- Resource

  Resources are provided by components and can be accessed by users, such as HDFS files or folders, queues in Yarn, and databases, tables, and columns in Hive.

- User

  A User is a visitor to the system. The rights of each user are obtained based on the policy associated with the user. Information about users, user groups, and roles in the LDAP is periodically synchronized to the Ranger.

- Permission

  In a policy, you can configure various access conditions for resources, such as file read and write, permission conditions, rejection conditions, and exception conditions.

9.12.1.2 Right Mechanism

FusionInsight adopts the Lightweight Directory Access Protocol (LDAP) to store data of users and user groups. Information about role definitions is stored in the relational database and the mapping between roles and rights is saved in components.

FusionInsight uses Kerberos for unified authentication.

The verification process of user rights is as follows:

1. A client (a user terminal or FusionInsight component service) invokes the FusionInsight authentication interface.
2. FusionInsight uses the login username and password for Kerberos authentication.
3. If the authentication succeeds, the client sends a request for accessing the server (a FusionInsight component service).
4. The server finds the user group and role to which the login user belongs.
5. The server obtains all rights of the user group and the role.
6. The server checks whether the client has the right to access the resources it applies for.

Example (RBAC):

There are three files in HDFS, that is, fileA, fileB, and fileC.

- roleA has read and write right for fileA, and roleB has the read right for fileB.
- groupA is bound to roleA, and groupB is bound to roleB.
- userA belongs to groupA and roleB, and userB belongs to groupB.

When userA successfully logs in to the system and accesses the HDFS:

1. HDFS obtains the role (roleB) to which userA is bound.
2. HDFS also obtains the role (roleA) to which the user group of userA is bound.
3. In this case, userA has all the rights of roleA and roleB.
4. As a result, userA has read and write rights for fileA, has the read right on fileB, and has no right for fileC.

Similarly, when userB successfully logs in to the system and accesses the HDFS:

1. userB only has the rights of roleB.
2. As a result, userB has the read right on fileB, and has no rights for fileA and fileC.

9.12.1.3 Authentication Policies

The big data platform performs user identity authentication to prevent invalid users from accessing the cluster. The cluster provides authentication capabilities in both security mode and normal mode.

Security Mode

The clusters in security mode use the Kerberos authentication protocol for security authentication. The Kerberos protocol supports mutual authentication between clients and servers. This eliminates the risks incurred by sending user credentials over the network for simulated authentication. In clusters, KrbServer provides the Kerberos authentication support.

Kerberos user object

In the Kerberos protocol, each user object is a principal. A complete principal consists of username and domain name. In O&M or application development scenarios, the user identity must be verified before a client connects to a server. Users for O&M and service operations are classified into human-machine and machine-machine users. The password of human-machine users is manually configured, while the password of machine-machine users is generated by the system randomly.

Kerberos authentication

Kerberos supports password and keytab authentication. The validity period of authentication is 24 hours by default.

- Password authentication: User identity is verified by entering the correct password. This mode mainly used in O&M scenarios where human-machine users are used. The configuration command is **kinit** *Username*.
- Keytab authentication: Keytab files contain users' principal and encrypted credential information. When keytab files are used for authentication, the system automatically uses encrypted credential information to perform authentication and the user password does not need to be entered. This mode is mainly used in component application development scenarios where machine-machine users are used. Keytab authentication can also be configured using the **kinit** command.

Normal Mode

Different components in a normal cluster use the native open-source authentication mode and do not support the **kinit** authentication command. FusionInsight Manager (including DBService, KrbServer, and LdapServer) uses the username and password for authentication. **Table 9-82** lists the authentication modes used by components.

Table 9-82 Component authentication modes

Service	Authentication Mode
ClickHouse	Simple authentication
Flume	No authentication
HBase	● Web UI: No authentication ● Client: simple authentication
HDFS	● Web UI: no authentication ● Client: simple authentication
Hive	Simple authentication
Hue	Username and password authentication
Kafka	No authentication
Loader	● Web UI: username and password authentication ● Client: no authentication
MapReduce	● Web UI: no authentication ● Client: no authentication

Service	Authentication Mode
Oozie	● Web UI: username and password authentication ● Client: simple authentication
Spark2x	● Web UI: no authentication ● Client: simple authentication
Storm	No authentication
YARN	● Web UI: no authentication ● Client: simple authentication
ZooKeeper	Simple authentication

The authentication modes are as follows:

- Simple authentication: When the client connects to the server, the client automatically authenticates the user (for example, the OS user **root** or **omm**) by default. The authentication is imperceptible to the administrator or service user, which does not require **kinit**.
- Username and password authentication: Use the username and password of human-machine users in the cluster for authentication.
- No authentication: Any user can access the server by default.

9.12.1.4 Permission Verification Policies

Security Mode

After a user is authenticated by the big data platform, the system determines whether to verify the user's permission based on the actual permission management configuration to ensure that the user has limited or all permission on resources. If the user does not have the permission for accessing cluster resources, the system administrator must grant the required permission to the user. Otherwise, the user fails to access the resources. The cluster provides permission verification capabilities in both security mode and normal mode. The specific permission items of the components are the same in the two modes.

By default, the Ranger service is installed and Ranger authentication is enabled for a newly installed cluster in security mode. You can set fine-grained security access policies for accessing component resources through the permission plug-in of the component. If Ranger authentication is not required, administrators can manually disable it on the service page. After Ranger authentication is disabled, the system continues to perform permission control based on the role model of FusionInsight Manager when accessing component resources.

In a cluster in security mode, the following components support Ranger authentication: HDFS, YARN, Kafka, Hive, HBase, Storm, Impala, and Spark2x.

For a cluster upgraded from an earlier version, Ranger authentication is not used by default when users access component resources. The administrator can manually enable Ranger authentication after installing Ranger.

By default, all components in the cluster of the security edition authenticate access. The authentication function cannot be disabled.

Normal Mode

Different components in a normal cluster use their own native open-source authentication behavior. **Table 9-83** lists detailed permission verification modes.

In a normal cluster, Ranger supports permission control on component resources based on OS users. The following components support Ranger authentication: HBase, HDFS, Hive, Spark2x, and YARN.

Table 9-83 Component permission verification modes in normal clusters

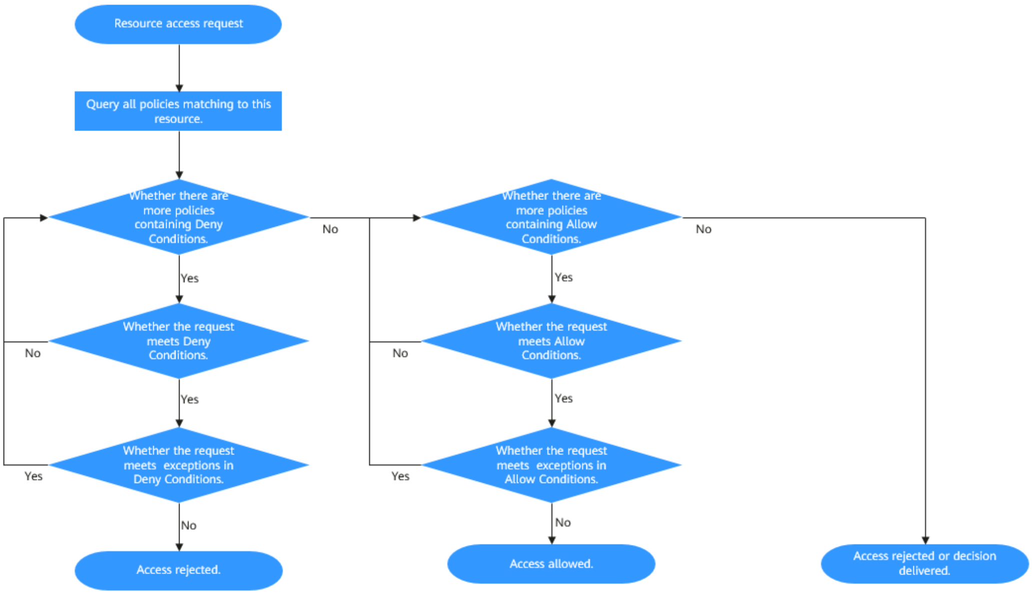
Service	Permission Verification	Permission Verification Enabling and Disabling
ClickHouse	Required	Not supported
Flume	Not required	Not supported
HBase	Not required	Supported
HDFS	Required	Supported
Hive	Not required	Not supported
Hue	Not required	Not supported
Kafka	Not required	Not supported
Loader	Not required	Not supported
MapReduce	Not required	Not supported
Oozie	Required	Not supported
Spark2x	Not required	Not supported
Storm	Not required	Not supported
YARN	Not required	Supported
ZooKeeper	Required	Supported

Condition Priorities of the Ranger Permission Policy

When configuring a permission policy for a resource, you can configure Allow Conditions, Exclude from Allow Conditions, Deny Conditions, and Exclude from Deny Conditions for the resource, to meet unexpected requirements in different scenarios.

The priorities of different conditions are listed in descending order: Exclude from Deny Conditions > Deny Conditions > Exclude from Allow Conditions > Allow Conditions

The following figure shows the process of determining condition priorities. If the component resource request does not match the permission policy in Ranger, the system rejects the access by default. However, for HDFS and Yarn, the system delivers the decision to the access control layer of the component for determination.


Resource access request
Query all policies matching to this resource.
Whether there are more policies containing Deny Conditions.
Yes
No
Whether the request meets Deny Conditions.
No
Yes
Whether the request meets exceptions in Deny Conditions.
Yes
No
Access rejected.
Whether there are more policies containing Allow Conditions.
Yes
No
Whether the request meets Allow Conditions.
No
Yes
Whether the request meets exceptions in Allow Conditions.
Yes
No
Access allowed.
Access rejected or decision delivered.

For example, if you want to grant the read and write permissions of the **FileA** folder to the **groupA** user group, but the user in the group is not **UserA**, you can add an allowed condition and an exception condition.

9.12.1.5 User Account List

User Classification

The MRS cluster provides the following three types of users. The system administrator needs to periodically change the passwords. It is not recommended to use the default passwords.

NOTE

This section describes the default users in the MRS cluster.

User Type	Description
System users	• User created on FusionInsight Manager for O&M and service scenarios. There are two types of users: – **Human-machine** user: used in scenarios such as FusionInsight Manager O&M and operations on a component client. When creating a user of this type, you need to set password and confirm password by referring to **Creating a User**. – **Machine-machine** user: used for system application development. • User who runs OMS processes

User Type	Description
Internal system users	Internal user to perform Kerberos authentication, process communications, save user group information, and associate user permissions. It is recommended that internal system users not be used in O&M scenarios. Operations can be performed as user **admin** or another user created by the system administrator based on service requirements.
Database users	● User who manages OMS database and accesses data ● User who runs service components (Hue, Hive, Loader, Oozie, Ranger, and DBService) in the database.

System Users

NOTE

- User **root** of the OS is required, the password of user **root** on all nodes must be the same.
- User **ldap** of the OS is required. Do not delete this account. Otherwise, the cluster may not work properly. The OS administrator maintains the password management policies.

User Type	Username	Initial Password	Description	Password Change Method
System administrator	admin	User-defined password	FusionInsight Manager administrator. **NOTE** By default, user **admin** does not have the management permission on other components. For example, when accessing the native UI of a component, the user fails to access the complete component information due to insufficient management permission on the component.	For details, see **Changing the Password for User admin**.

User Type	Username	Initial Password	Description	Password Change Method
Node OS user	ommdba	Random password	User that creates the system database. This user is an OS user generated on the management node and does not require a unified password. This account cannot be used for remote login.	For details, see **Changing the Password for an OS User**.
	omm	Bigdata123@	Internal running user of the system. This user is an OS user generated on all nodes and does not require a unified password.	

Internal System Users

User Type	Default User	Initial Password	Description	Password Change Method
Kerberos administrator	kadmin/admin	Admin@123	Used to add, delete, modify, and query user accounts on Kerberos.	For details, see **Changing the Password for the Kerberos Administrator**.
OMS Kerberos administrator	kadmin/admin	Admin@123	Used to add, delete, modify, and query user accounts on OMS Kerberos.	For details, see **Changing the Password for the OMS Kerberos Administrator**.

User Type	Default User	Initial Password	Description	Password Change Method
LDAP administrator	cn=root,dc=hadoop,dc=com	• Versions earlier than MRS 3.1.2: LdapChangeMe@123 • MRS 3.1.2 or later: randomly generated by the system	Used to add, delete, modify, and query the user account information on LDAP.	• For versions earlier than MRS 3.1.2, see **Changing the Passwords of the LDAP Administrator and the LDAP User (Including OMS LDAP)**. • For MRS 3.1.2 or later, see **Modifying OMS Service Configuration Parameters**.
OMS LDAP administrator	cn=root,dc=hadoop,dc=com	• Versions earlier than MRS 3.1.2: LdapChangeMe@123 • MRS 3.1.2 or later: randomly generated by the system	Used to add, delete, modify, and query the user account information on OMS LDAP.	

User Type	Default User	Initial Password	Description	Password Change Method
LDAP user	cn=pg_search_dn,ou=Users,dc=hadoop,dc=com	Randomly generated by the system	Used to query information about users and user groups on LDAP.	
OMS LDAP user	cn=pg_search_dn,ou=Users,dc=hadoop,dc=com	Randomly generated by the system	Used to query information about users and user groups on OMS LDAP.	
LDAP administrator account	cn=krbkdc,ou=Users,dc=hadoop,dc=com	● Versions earlier than MRS 3.1.2: LdapChangeMe@123 ● MRS 3.1.2 or later: randomly generated by the system	Used to query Kerberos component authentication account information.	● For versions earlier than MRS 3.1.2, see **Changing the Password for the LDAP Administrator**. ● For MRS 3.1.2 or later, see **Modifying OMS Service Configuration Parameters**.

User Type	Default User	Initial Password	Description	Password Change Method
	cn=krbadmin,ou=Users,dc=hadoop,dc=com	● Versions earlier than MRS 3.1.2: LdapChangeMe@123 ● MRS 3.1.2 or later: randomly generated by the system	Used to add, delete, modify, and query Kerberos component authentication account information.	

User Type	Default User	Initial Password	Description	Password Change Method
Component running user	hdfs	Hdfs@123	This user is the HDFS system administrator and has the following permissions: 1. File system operation permissions: - Views, modifies, and creates files. - Views and creates directories. - Views and modifies the groups where files belong. - Views and sets disk quotas for users. 2. HDFS management operation permissions: - Views the web UI status. - Views and sets the active and standby HDFS status. - Enters and exits the HDFS in security mode. - Checks the HDFS file system. 3. Logs in to the FTP service page.	For details, see **Changing the Password for a Component Running User**.

User Type	Default User	Initial Password	Description	Password Change Method
	hbase	Hbase@123	This user is the HBase and HBase1 to HBase4 system administrator and has the following permissions: ● Cluster management permission: Performs **Enable** and **Disable** operations on tables to trigger MajorCompact and ACL operations. ● Grants and revokes permissions, and shuts down the cluster. ● Table management permission: Creates, modifies, and deletes tables. ● Data management permission: Reads data in tables, column families, and columns. ● Logs in to the HMaster web UI. ● Logs in to the FTP service page.	

User Type	Default User	Initial Password	Description	Password Change Method
	mapred	Mapred@123	This user is the MapReduce system administrator and has the following permissions: ● Submits, stops, and views the MapReduce tasks. ● Modifies the Yarn configuration parameters. ● Logs in to the FTP service page. ● Logs in to the Yarn web UI.	
	zookeeper	ZooKeeper@123	This user is the ZooKeeper system administrator and has the following permissions: ● Adds, deletes, modifies, and queries all nodes in ZooKeeper. ● Modifies and queries quotas of all nodes in ZooKeeper.	
	rangeradmin	Rangeradmin@123	This user has the Ranger system management permissions and user permissions: ● Ranger web UI management permission ● Management permission of each component that uses Ranger authentication	
	rangerauditor	Rangerauditor@123	Default audit user of the Ranger system.	

User Type	Default User	Initial Password	Description	Password Change Method
	hive	Hive@123	This user is the Hive system administrator and has the following permissions: 1. Hive administrator permissions: ● Creates, deletes, and modifies a database. ● Creates, queries, modifies, and deletes a table. ● Queries, inserts, and uploads data. 2. HDFS file operation permissions: ● Views, modifies, and creates files. ● Views and creates directories. ● Views and modifies the groups where files belong. 3. Submits and stops the MapReduce tasks. 4. Ranger policy management permission	

User Type	Default User	Initial Password	Description	Password Change Method
	hive1	Hive1@123	This user is the Hive1 system administrator and has the following permissions: 1. Hive1 administrator permissions: - Creates, deletes, and modifies a database. - Creates, queries, modifies, and deletes a table. - Queries, inserts, and uploads data. 2. HDFS file operation permissions: - Views, modifies, and creates files. - Views and creates directories. - Views and modifies the groups where files belong. 3. Submits and stops the MapReduce tasks. 4. Ranger policy management permission	

User Type	Default User	Initial Password	Description	Password Change Method
	hive2	Hive2@123	This user is the Hive2 system administrator and has the following permissions: 1. Hive2 administrator permissions: ● Creates, deletes, and modifies a database. ● Creates, queries, modifies, and deletes a table. ● Queries, inserts, and uploads data. 2. HDFS file operation permissions: ● Views, modifies, and creates files. ● Views and creates directories. ● Views and modifies the groups where files belong. 3. Submits and stops the MapReduce tasks. 4. Ranger policy management permission	

User Type	Default User	Initial Password	Description	Password Change Method
	hive3	Hive3@123	This user is the Hive3 system administrator and has the following permissions: 1. Hive3 administrator permissions: ● Creates, deletes, and modifies a database. ● Creates, queries, modifies, and deletes a table. ● Queries, inserts, and uploads data. 2. HDFS file operation permissions: ● Views, modifies, and creates files. ● Views and creates directories. ● Views and modifies the groups where files belong. 3. Submits and stops the MapReduce tasks. 4. Ranger policy management permission	

User Type	Default User	Initial Password	Description	Password Change Method
	hive4	Hive4@123	This user is the Hive4 system administrator and has the following permissions: 1. Hive4 administrator permissions: - Creates, deletes, and modifies a database. - Creates, queries, modifies, and deletes a table. - Queries, inserts, and uploads data. 2. HDFS file operation permissions: - Views, modifies, and creates files. - Views and creates directories. - Views and modifies the groups where files belong. 3. Submits and stops the MapReduce tasks. 4. Ranger policy management permission	

User Type	Default User	Initial Password	Description	Password Change Method
	kafka	Kafka@123	This user is the Kafka system administrator and has the following permissions: ● Creates, deletes, produces, and consumes the topic; modifies the topic configuration. ● Controls the cluster metadata, modifies the configuration, migrates the replica, elects the leader, and manages ACL. ● Submits, queries, and deletes the consumer group offset. ● Queries the delegation token. ● Queries and submits the transaction.	
	storm	Admin@123	Storm system administrator User permission: Submits Storm tasks.	
	rangerusersync	Randomly generated by the system	Synchronizes users and internal users of user groups.	
	rangertagsync	Randomly generated by the system	Internal user for synchronizing tags.	

User Type	Default User	Initial Password	Description	Password Change Method
	oms/manager	Randomly generated by the system	Controller and NodeAgent authentication user. The user has the permission on the **supergroup** group.	
	backup/manager	Randomly generated by the system	User for running backup and restoration tasks. The user has the permission on the **supergroup**, **wheel**, and **ficommon** groups. After cross-system mutual trust is configured, the user has the permission to access data in the HDFS, HBase, Hive, and ZooKeeper systems.	

User Type	Default User	Initial Password	Description	Password Change Method
	hdfs/hadoop.*<System domain name>*	Randomly generated by the system	This user is used to start the HDFS and has the following permissions: 1. File system operation permissions: - Views, modifies, and creates files. - Views and creates directories. - Views and modifies the groups where files belong. - Views and sets disk quotas for users. 2. HDFS management operation permissions: - Views the web UI status. - Views and sets the active and standby HDFS status. - Enters and exits the HDFS in security mode. - Checks the HDFS file system. 3. Logs in to the FTP service page.	

User Type	Default User	Initial Password	Description	Password Change Method
	mapred/hadoop.*<System domain name>*	Randomly generated by the system	This user is used to start the MapReduce and has the following permissions: ● Submits, stops, and views the MapReduce tasks. ● Modifies the Yarn configuration parameters. ● Logs in to the FTP service page. ● Logs in to the Yarn web UI.	
	mr_zk/hadoop.*<System domain name>*	Randomly generated by the system	Used for MapReduce to access ZooKeeper.	
	hbase/hadoop.*<System domain name>*	Randomly generated by the system	User for the authentication between internal components during the HBase system startup.	
	hbase/zkclient.*<System domain name>*	Randomly generated by the system	User for HBase to perform ZooKeeper authentication in a security mode cluster.	
	thrift/hadoop.*<System domain name>*	Randomly generated by the system	ThriftServer system startup user.	

User Type	Default User	Initial Password	Description	Password Change Method
	thrift/*<hostname>*	Randomly generated by the system	User for the ThriftServer system to access HBase. This user has the read, write, execution, creation, and administration permission on all NameSpaces and tables of HBase. *<hostname>* indicates the name of the host where the ThriftServer node is installed in the cluster.	

User Type	Default User	Initial Password	Description	Password Change Method
	hive/hadoop.*<System domain name>*	Randomly generated by the system	User for the authentication between internal components during the Hive system startup. The user permissions are as follows: 1. Hive administrator permissions: ● Creates, deletes, and modifies a database. ● Creates, queries, modifies, and deletes a table. ● Queries, inserts, and uploads data. 2. HDFS file operation permissions: ● Views, modifies, and creates files. ● Views and creates directories. ● Views and modifies the groups where files belong. 3. Submits and stops the MapReduce tasks.	

User Type	Default User	Initial Password	Description	Password Change Method
	hive1/ hadoop.*<System domain name>*	Randomly generated by the system	User for the authentication between internal components during the Hive1 system startup. The user permissions are as follows: 1. Hive1 administrator permissions: - Creates, deletes, and modifies a database. - Creates, queries, modifies, and deletes a table. - Queries, inserts, and uploads data. 2. HDFS file operation permissions: - Views, modifies, and creates files. - Views and creates directories. - Views and modifies the groups where files belong. 3. Submits and stops the MapReduce tasks.	

User Type	Default User	Initial Password	Description	Password Change Method
	hive2/hadoop.*<System domain name>*	Randomly generated by the system	User for the authentication between internal components during the Hive2 system startup. The user permissions are as follows: 1. Hive2 administrator permissions: ● Creates, deletes, and modifies a database. ● Creates, queries, modifies, and deletes a table. ● Queries, inserts, and uploads data. 2. HDFS file operation permissions: ● Views, modifies, and creates files. ● Views and creates directories. ● Views and modifies the groups where files belong. 3. Submits and stops the MapReduce tasks.	

User Type	Default User	Initial Password	Description	Password Change Method
	hive3/hadoop.*<System domain name>*	Randomly generated by the system	User for the authentication between internal components during the Hive3 system startup. The user permissions are as follows: 1. Hive3 administrator permissions: ● Creates, deletes, and modifies a database. ● Creates, queries, modifies, and deletes a table. ● Queries, inserts, and uploads data. 2. HDFS file operation permissions: ● Views, modifies, and creates files. ● Views and creates directories. ● Views and modifies the groups where files belong. 3. Submits and stops the MapReduce tasks.	

User Type	Default User	Initial Password	Description	Password Change Method
	hive4/ hadoop.*<System domain name>*	Randomly generated by the system	User for the authentication between internal components during the Hive4 system startup. The user permissions are as follows: 1. Hive4 administrator permissions: - Creates, deletes, and modifies a database. - Creates, queries, modifies, and deletes a table. - Queries, inserts, and uploads data. 2. HDFS file operation permissions: - Views, modifies, and creates files. - Views and creates directories. - Views and modifies the groups where files belong. 3. Submits and stops the MapReduce tasks.	
	loader/ hadoop.*<System domain name>*	Randomly generated by the system	User for Loader system startup and Kerberos authentication	

User Type	Default User	Initial Password	Description	Password Change Method
	HTTP/*<hostname>*	Randomly generated by the system	Used to connect to the HTTP interface of each component. *<hostname>* indicates the host name of a node in the cluster.	
	hue	Randomly generated by the system	User for Hue system startup, Kerberos authentication, and HDFS and Hive access	
	flume	Randomly generated by the system	User for Flume system startup and HDFS and Kafka access. The user has read and write permission of the HDFS directory **/flume**.	
	flume_server	Randomly generated by the system	User for Flume system startup and HDFS and Kafka access. The user has read and write permission of the HDFS directory **/flume**.	
	spark2x/hadoop.*<System domain name>*	Randomly generated by the system	This user is the Spark2x system administrator and has the following user permissions: 1. Starts the Spark2x service. 2. Submits Spark2x tasks.	
	spark_zk/hadoop.*<System domain name>*	Randomly generated by the system	Used for Spark2x to access ZooKeeper.	

User Type	Default User	Initial Password	Description	Password Change Method
	spark2x1/ hadoop.*<System domain name>*	Randomly generated by the system	This user is the Spark2x1 system administrator and has the following user permissions: 1. Starts the Spark2x1 service. 2. Submits Spark2x tasks.	
	spark2x2/ hadoop.*<System domain name>*	Randomly generated by the system	This user is the Spark2x2 system administrator and has the following user permissions: 1. Starts the Spark2x2 service. 2. Submits Spark2x tasks.	
	spark2x3/ hadoop.*<System domain name>*	Randomly generated by the system	This user is the Spark2x3 system administrator and has the following user permissions: 1. Starts the Spark2x3 service. 2. Submits Spark2x tasks.	
	spark2x4/ hadoop.*<System domain name>*	Randomly generated by the system	This user is the Spark2x4 system administrator and has the following user permissions: 1. Starts the Spark2x4 service. 2. Submits Spark2x tasks.	
	zookeeper/ hadoop.*<System domain name>*	Randomly generated by the system	ZooKeeper system startup user.	

User Type	Default User	Initial Password	Description	Password Change Method
	zkcli/hadoop.*<System domain name>*	Randomly generated by the system	ZooKeeper server login user.	
	oozie	Randomly generated by the system	User for Oozie system startup and Kerberos authentication.	
	kafka/hadoop.*<System domain name>*	Randomly generated by the system	Used for security authentication of Kafka.	
	storm/hadoop.*<System domain name>*	Randomly generated by the system	Storm system startup user.	
	storm_zk/hadoop.*<System domain name>*	Randomly generated by the system	Used for the Worker process to access ZooKeeper.	
	flink/hadoop.*<System domain name>*	Randomly generated by the system	Internal user of the Flink service.	
	check_ker_M	Randomly generated by the system	User who performs a system internal test about whether the Kerberos service is normal.	
	clickhouse/hadoop.*<System domain name>*	Randomly generated by the system	Used for security authentication of ClickHouse. This user is an internal user and can be used only in the cluster.	

User Type	Default User	Initial Password	Description	Password Change Method
	default	None	ClickHouse internal user, which is an administrator user that can be used only in non-security mode.	
	rangeradmin/hadoop.*<System domain name>*	Randomly generated by the system	Ranger system startup user, which is used for authentication between internal components.	
	tez	Randomly generated by the system	User for TezUI system startup, Kerberos authentication, and access to Yarn	
	K/M	Randomly generated by the system	Kerberos internal functional user. This user cannot be deleted, and its password cannot be changed. This internal account can only be used on nodes where Kerberos service is installed.	None
	kadmin/changepw	Randomly generated by the system		
	kadmin/history	Randomly generated by the system		
	krbtgt*<System domain name>*	Randomly generated by the system		

User Type	Default User	Initial Password	Description	Password Change Method
LDAP user	admin	None	FusionInsight Manager administrator. The primary group is **compcommon**, which does not have the group permission but has the permission of the **Manager_administrator** role.	The LDAP user cannot log in to the system, and the password cannot be changed.
	backup		The primary group is **compcommon**.	
	backup/manager		The primary group is **compcommon**.	
	oms		The primary group is **compcommon**.	
	oms/manager		The primary group is **compcommon**.	
	clientregister		The primary group is **compcommon**.	
	zookeeper		The primary group is **hadoop**.	
	zookeeper/hadoop.*<System domain name>*		The primary group is **hadoop**.	
	zkcli		The primary group is **hadoop**.	
	zkcli/hadoop.<*System domain name*>		The primary group is **hadoop**.	
	flume		The primary group is **hadoop**.	
	flume_server		The primary group is **hadoop**.	

User Type	Default User	Initial Password	Description	Password Change Method
	hdfs		The primary group is **hadoop**.	
	hdfs/hadoop.*<System domain name>*		The primary group is **hadoop**.	
	mapred		The primary group is **hadoop**.	
	mapred/hadoop.*<System domain name>*		The primary group is **hadoop**.	
	mr_zk		The primary group is **hadoop**.	
	mr_zk/hadoop.*<System domain name>*		The primary group is **hadoop**.	
	hue		The primary group is **supergroup**.	
	hive		The primary group is **hive**.	
	hive/hadoop.*<System domain name>*		The primary group is **hive**.	
	hive1		The primary group is **hive1**.	
	hive1/hadoop.*<System domain name>*		The primary group is **hive1**.	
	hive2		The primary group is **hive2**.	

User Type	Default User	Initial Password	Description	Password Change Method
	hive2/ hadoop.*<System domain name>*		The primary group is **hive2**.	
	hive3		The primary group is **hive3**.	
	hive3/ hadoop.*<System domain name>*		The primary group is **hive3**.	
	hive4		The primary group is **hive4**.	
	hive4/ hadoop.*<System domain name>*		The primary group is **hive4**.	
	hbase		The primary group is **hadoop**.	
	hbase/ hadoop.*<System domain name>*		The primary group is **hadoop**.	
	thrift		The primary group is **hadoop**.	
	thrift/ hadoop.*<System domain name>*		The primary group is **hadoop**.	
	oozie		The primary group is **hadoop**.	
	hbase/ zkclient.*<System domain name>*		The primary group is **hadoop**.	

User Type	Default User	Initial Password	Description	Password Change Method
	loader		The primary group is **hadoop**.	
	loader/ hadoop.*<System domain name>*		The primary group is **hadoop**.	
	spark2x		The primary group is **hadoop**.	
	spark2x/ hadoop.*<System domain name>*		The primary group is **hadoop**.	
	spark_zk		The primary group is **hadoop**.	
	spark2x1		The primary group is **hadoop**.	
	spark2x1/ hadoop.*<System domain name>*		The primary group is **hadoop**.	
	spark2x2		The primary group is **hadoop**.	
	spark2x2/ hadoop.*<System domain name<*		The primary group is **hadoop**.	
	spark2x3		The primary group is **hadoop**.	
	spark2x3/ hadoop.*<System domain name>*		The primary group is **hadoop**.	
	spark2x4		The primary group is **hadoop**.	

User Type	Default User	Initial Password	Description	Password Change Method
	spark2x4/hadoop.*<System domain name>*		The primary group is **hadoop**.	
	kafka		The primary group is **kafkaadmin**.	
	kafka/hadoop.*<System domain name>*		The primary group is **kafkaadmin**.	
	storm		The primary group is **stormadmin**.	
	storm/hadoop.*<System domain name>*		The primary group is **stormadmin**.	
	storm_zk		The primary group is **storm**.	
	storm_zk/hadoop.*<System domain name>*		The primary group is **storm**.	
	kms/hadoop		The primary group is **kmsadmin**.	
	knox		The primary group is **compcommon**.	
	executor		The primary group is **compcommon**.	
	rangeradmin		The primary group is **supergroup**.	
	rangeradmin/hadoop.*<System domain name>*		The primary group is **supergroup**.	

User Type	Default User	Initial Password	Description	Password Change Method
	rangerusersync		The primary group is **supergroup**.	
	rangertagsync		The primary group is **supergroup**.	
	rangerauditor		The primary group is **compcommon**.	

NOTE

Log in to FusionInsight Manager, choose **System** > **Permission** > **Domain and Mutual Trust**, and check the value of **Local Domain**. In the preceding table, all letters in the system domain name contained in the username of the system internal user are lowercase letters.

For example, if **Local Domain** is set to **9427068F-6EFA-4833-B43E-60CB641E5B6C.COM**, the username of default HDFS startup user is **hdfs/hadoop.9427068f-6efa-4833-b43e-60cb641e5b6c.com**.

Database Users

The system database users include OMS database users and DBService database users.

Database Type	Default User	Initial Password	Description	Password Change Method
OMS database	ommdba	dbChangeMe@123456	OMS database administrator who performs maintenance operations, such as creating, starting, and stopping.	For details, see **Changing the Password of the OMS Database Administrator**.
	omm	ChangeMe@123456	User for accessing OMS database data	For details, see **Changing the Password for the Data Access User of the OMS Database**.

Database Type	Default User	Initial Password	Description	Password Change Method
DBService database	omm	dbserverAdmin@123	Administrator of the GaussDB database in the DBService component	- For versions earlier than MRS 3.1.2, see **Changing the Password for a Component Database User**. - For MRS 3.1.2 or later, see **Changing the Password for User omm in DBService**.
	compdbuser	Random password	MRS 3.1.2 or later: Administrator of the GaussDB database in the DBService component. It is used in service O&M scenarios. If the password of this account has expired, you need to reset the password upon your first login.	For details, see **Changing the Password for User compdbuser of the DBService Database**.

Database Type	Default User	Initial Password	Description	Password Change Method
	hive	- Versions earlier than MRS 3.1.2: HiveUser@ - MRS 3.1.2 or later: ran	User for Hive to connect to the DBService database **hivemeta**.	- For versions earlier than MRS 3.1.2, see **Changing the Password for a Component Database User**. - For MRS 3.1.2 or later, see **Resetting the Component Database User Password**.

Database Type	Default User	Initial Password	Description	Password Change Method
		dompassword		

Database Type	Default User	Initial Password	Description	Password Change Method
	hive1	● Versions earlier than MRS 3.1.2: HiveUser@ ● MRS 3.1.2 or later: ran	User for Hive1 to connect to the DBService database **hivemeta1**.	

Database Type	Default User	Initial Password	Description	Password Change Method
		dompassword		

Database Type	Default User	Initial Password	Description	Password Change Method
	hive2	● Versions earlier than MRS 3.1.2: HiveUser@ ● MRS 3.1.2 or later: ran	User for Hive2 to connect to the DBService database **hivemeta2**.	

Database Type	Default User	Initial Password	Description	Password Change Method
		dompassword		

Database Type	Default User	Initial Password	Description	Password Change Method
	hive3	● Versions earlier than MRS 3.1.2: HiveUser@ ● MRS 3.1.2 or later: ran	User for Hive3 to connect to the DBService database **hivemeta3**.	

Database Type	Default User	Initial Password	Description	Password Change Method
		dompassword		

Database Type	Default User	Initial Password	Description	Password Change Method
	hive4	● Versions earlier than MRS 3.1.2: HiveUser@ ● MRS 3.1.2 or later: ran	User for Hive4 to connect to the DBService database **hivemeta4**.	

Database Type	Default User	Initial Password	Description	Password Change Method
		dompassword		

Database Type	Default User	Initial Password	Description	Password Change Method
	hive*NN*	● Versions earlier than MRS 3.1.2: HiveUser@ ● MRS 3.1.2 or later: ran	User for **Hive-***N* to connect to the DBService database **hive***N***meta** when multiple services are installed. For example, the user for **Hive-1** to connect to the DBService database **hive1meta** is **hive11**.	

Database Type	Default User	Initial Password	Description	Password Change Method
		dompassword		

Database Type	Default User	Initial Password	Description	Password Change Method
	hue	● Versions earlier than MRS 3.1.2: Hue User@123 ● MRS 3.1.2 or later:	User for Hue to connect to the DBService database **hue**.	

Database Type	Default User	Initial Password	Description	Password Change Method
		random password		

Database Type	Default User	Initial Password	Description	Password Change Method
	sqoop	● Versions earlier than MRS 3.1.2: SqoopUser@ ● MRS 3.1.2 or later: r	User for Loader to connect to the DBService database **sqoop**.	

Database Type	Default User	Initial Password	Description	Password Change Method
		andompassword		

Database Type	Default User	Initial Password	Description	Password Change Method
	sqoop*N*	● Versions earlier than MRS 3.1.2: SqoopUser@ ● MRS 3.1.2 or later: r	User for **Loader-*N*** to connect to the DBService database **sqoop*N*** when multiple services are installed. For example, the user for **Loader-1** to connect to the DBService database **sqoop1** is **sqoop1**.	

Database Type	Default User	Initial Password	Description	Password Change Method
		andom password		

Database Type	Default User	Initial Password	Description	Password Change Method
	oozie	- Versions earlier than MRS 3.1.2: OozieUser@ - MRS 3.1.2 or later: ra	User for Oozie to connect to the DBService database **oozie**.	

Database Type	Default User	Initial Password	Description	Password Change Method
		ndompassword		

Database Type	Default User	Initial Password	Description	Password Change Method
	oozie*N*	• Versions earlier than MRS 3.1.2: OozieUser@ • MRS 3.1.2 or later: ra	User for **Oozie-***N* to connect to the DBService database **oozie***N* when multiple services are installed. For example, the user for **Oozie-1** to connect to the DBService database **oozie1** is **oozie1**.	

Database Type	Default User	Initial Password	Description	Password Change Method
		ndompassword		

Database Type	Default User	Initial Password	Description	Password Change Method
	rangeradmin	● Versions earlier than MRS 3.1.2: OozieUser@ ● MRS 3.1.2 or later: ra	User for Ranger to connect to the DBService database.	

Database Type	Default User	Initial Password	Description	Password Change Method
		ndompassword		
	kafkaui	Random password	User for Kafka UI to connect to the DBService database. This user exists only in MRS 3.1.2 or later.	
	flink	Random password	User for Flink to connect to the DBService database. This user exists only in MRS 3.1.2 or later.	

9.12.1.6 Default Permission Information

Role

Default Role	Description
Manager_administrator	Manager administrator who has all permissions for Manager. Manager administrators can create first-level tenants, create and modify user groups, and specify user permissions.
Manager_operator	Manager operator who has all the permissions on the **Homepage**, **Cluster**, **Hosts**, and **O&M** tab pages.
Manager_auditor	Manager auditor who has all permissions on the **Audit** tab page. Manager auditors can view and manage Manager system audit logs.

Default Role	Description
Manager_viewer	Manager viewer who has the permission to view information about **Homepage**, **Cluster**, **Hosts**, **Alarm**, **Events**, and **System > Permission**.
Manager_tenant	Manager tenant administrator. This role can create and manage sub-tenants for the non-leaf tenants to which the current user belongs. It has the permission to view alarms and events on **O&M > Alarm**.
System_administrator	System administrator, this role has Manager system administrator rights and all services administrator rights.
default	This role is the default role created for the **default** tenant. It has the management permissions on the Yarn component and the default queue. The default role of the default tenant that is not the first cluster to be installed is **c***<cluster ID>***_default**.
Manager_administrator_180	FusionInsight Manager System administrator group. Internal system user group, which is used only between components.
Manager_auditor_181	FusionInsight Manager system auditor group. Internal system user group, which is used only between components.
Manager_operator_182	FusionInsight Manager system operator group. Internal system user group, which is used only between components.
Manager_viewer_183	FusionInsight Manager system viewer group. Internal system user group, which is used only between components.
System_administrator_186	System administrator group. Internal system user group, which is used only between components.
Manager_tenant_187	Tenant system user group. Internal system user group, which is used only between components.
default_1000	This group is created for tenant. Internal system user group, which is used only between components.

User group

Type	Default User Group	Description
OS User Group	hadoop	Users added to this group are granted the permission to submit all Yarn queue tasks.
	hadoopmanager	Users added to this user group can have the O&M manager rights of HDFS and Yarn. The O&M manager of HDFS can access the NameNode WebUI and perform active to standby switchover manually. The O&M manager of Yarn can access the ResourceManager WebUI, operate NodeManager nodes, refresh queues, and set node labels, but cannot submit tasks.
	hetuadmin	HetuEngine administrator group. Users in this group have the permission to perform operations on HSConsole.
	hive	Common user group. Hive users must belong to this user group.
	hive1	Common user group. Hive1 users must belong to this user group.
	hive2	Common user group. Hive2 users must belong to this user group.
	hive3	Common user group. Hive3 users must belong to this user group.
	hive4	Common user group. Hive4 users must belong to this user group.
	kafka	Kafka common user group. A user in this group can access a topic only when a user in the kafkaadmin group grants the read and write permission of the topic to the user.
	kafkaadmin	Kafka administrator group. Users in this group have the rights to create, delete, authorize, read, and write all topics.
	kafkasuperuser	Topic read/write user group of Kafka. Users added to this group have the read and write permissions on all topics.
	storm	Users who are added to the storm user group can submit topologies and manage their own topologies.
	stormadmin	Users who are added to the stormadmin user group can have the storm administrator rights and can submit topologies and manage all topologies.
	supergroup	Users added to this user group can have the administrator rights of HBase, HDFS and Yarn and can use Hive.
	yarnviewgroup	Indicates the read-only user group of the Yarn task. Users in this user group can have the view permission on Yarn and MapReduce tasks.

Type	Default User Group	Description
	check_sec_ldap	Perform internal test on the active LDAP to see whether it works properly. This user group is generated randomly in a test and automatically deleted after the test is complete. Internal system user group, which is used only between components.
	compcommon	System internal group for accessing cluster system resources. All system users and system running users are added to this user group by default.
OS User Group	wheel	Primary group of the FusionInsight internal running user omm.
	ficommon	System common group that corresponds to **compcommon** for accessing cluster common resource files stored in the OS.

NOTE

If the current cluster is not the cluster that is installed for the first time in FusionInsight Manager, the default user group name of all components except Manager in the cluster is **c***<cluster ID>_ default user group name*, for example, **c2_hadoop**.

User

For details, see **User Account List**.

Service-related User Security Parameters

- **HDFS**

  The **dfs.permissions.superusergroup** parameter specifies the administrator group with the highest permission on the HDFS. The default value is **supergroup**.

- **Spark2x and Corresponding Multi-Instances**

  The **spark.admin.acls** parameter specifies the administrator list of the Spark2x. Members in the list are authorized to manage all Spark tasks. Users not added in the list cannot manage all Spark tasks. The default value is **admin**.

9.12.1.7 FusionInsight Manager Security Functions

You can query and set user rights data through the following FusionInsight Manager modules:

- User management: Users can be added, deleted, modified, queried, bound to user groups, and assigned with roles.

  For details, see **Managing Users**.

- User group management: User groups can be added, deleted, modified, queried, and bound to roles.

  For details, see **Managing User Groups**.
- Role management: Roles can be added, deleted, modified, queried, and assigned with the resource access rights of one or multiple components.

  For details, see **Managing Roles**.
- Tenant management: Tenants can be added, deleted, modified, queried, and bound to component resources. FusionInsight generates a role for each tenant to facilitate management. If a tenant is assigned with the rights of some resources, its corresponding role also has these rights.

  For details, see **Tenant Resources**.

9.12.2 Account Management

9.12.2.1 Account Security Settings

9.12.2.1.1 Unlocking LDAP Users and Management Accounts

Scenario

If the LDAP user **cn=pg_search_dn,ou=Users,dc=hadoop,dc=com** and LDAP management accounts **cn=krbkdc,ou=Users,dc=hadoop,dc=com** and **cn=krbadmin,ou=Users,dc=hadoop,dc=com** are locked, the administrator must unlock these accounts.

NOTE

If you input an incorrect password for the LDAP user or management account for five consecutive times, the LDAP user or management account is locked. The account is automatically unlocked after 5 minutes.

Procedure

Step 1 Log in to the active management node as user **omm**.

Step 2 Run the following command to go to the related directory:

cd ${BIGDATA_HOME}/om-server/om/ldapserver/ldapserver/local/script

Step 3 Run the following command to unlock the LDAP user or management account:

./ldapserver_unlockUsers.sh *USER_NAME*

In the command, *USER_NAME* indicates the name of the user to be unlocked.

For example, to unlock the LDAP management **account cn=krbkdc,ou=Users,dc=hadoop,dc=com**, run the following command:

./ldapserver_unlockUsers.sh krbkdc

After the script is executed, enter the password of user **krbkdc** after **ROOT_DN_PASSWORD**. If the following information is displayed, the account is successfully unlocked.

```
Unlock user krbkdc successfully.
```

----End

9.12.2.1.2 Internal an Internal System User

Scenario

If the service is abnormal, the internal user of the system may be locked. Unlock the user promptly, or the cluster cannot run properly. For the list of system internal users, see **User Account List** in . The internal user of the system cannot be unlocked using FusionInsight Manager.

Prerequisites

Obtain the default password of the LDAP administrator **cn=root,dc=hadoop,dc=com** by referring to **User Account List** in .

Procedure

Step 1 Use the following method to confirm whether the internal system username is locked:

1. OLdap port number obtaining method:
   a. Log in to FusionInsight Manager, choose **System** > **OMS** > **oldap** > **Modify Configuration**.
   b. The **LDAP Listening Port** parameter value is **oldap port**.
2. Domain name obtaining method:
   a. Log in to FusionInsight Manager, choose **System** > **Permission** > **Domain and Mutual Trust**.
   b. The **Local Domain** parameter value is the domain name.

      For example, the domain name of the current system is **9427068F-6EFA-4833-B43E-60CB641E5B6C.COM**.
3. Run the following command on each node in the cluster as user **omm** to query the number of password authentication failures:

   ldapsearch -H ldaps://*OMS Floating IP Address*:*OLdap port* **-LLL -x -D cn=root,dc=hadoop,dc=com -b krbPrincipalName=***Internal system username***@***Domain name***,cn=***Domain name***,cn=krbcontainer,dc=hadoop,dc=com -w** *Password of LDAP administrator* **-e ppolicy | grep krbLoginFailedCount**

   For example, run the following command to check the number of password authentication failures for user **oms/manager**:

   ldapsearch -H ldaps://10.5.146.118:21750 -LLL -x -D cn=root,dc=hadoop,dc=com -b krbPrincipalName=oms/manager@9427068F-6EFA-4833-B43E-60CB641E5B6C.COM,cn=9427068F-6EFA-4833-B43E-60CB641E5B6C.COM,cn=krbcontainer,dc=hadoop,dc=com -w *Password of user cn=root,dc=hadoop,dc=com* **-e ppolicy | grep krbLoginFailedCount**

   ```
   krbLoginFailedCount: 5
   ```

4. Log in to FusionInsight Manager, choose **System** > **Permission** > **Security Policy** > **Password Policy**.
5. Check the value of the **Password Retries** parameter. If the value is less than or equal to the value of **krbLoginFailedCount**, the user is locked.

   NOTE

   You can also check whether internal users are locked by viewing operations logs.

Step 2 Log in to the active management node as user **omm** and run the following command to unlock the user:

sh ${BIGDATA_HOME}/om-server/om/share/om/acs/config/unlockuser.sh --userName *Internal system username*

Example: **sh ${BIGDATA_HOME}/om-server/om/share/om/acs/config/unlockuser.sh --userName oms/manager**

----End

9.12.2.1.3 Enabling and Disabling Permission Verification on Cluster Components

Scenario

HDFS and ZooKeeper verify the permission of users who attempt to access the services in both security and normal clusters by default. Users without related permission cannot access resources in HDFS and ZooKeeper. When the cluster is deployed in normal mode, HBase and YARN do not verify the permission of users who attempt to access the services by default. All users can access resources in HBase and YARN.

Based on actual service requirements, administrators can enable permission verification on HBase and YARN or disable permission verification on HDFS and ZooKeeper in normal clusters.

Impact on the System

After the enabling and disabling operations, the service configuration will expire. You need to restart the corresponding service for the configuration to take effect.

Enabling Permission Verification on HBase

Step 1 Log in to FusionInsight Manager.

Step 2 Click **Cluster**, click the name of the desired cluster, choose **Services** > **Ranger**, and click **Configurations**.

Step 3 Click **All Configurations**.

Step 4 Search for parameters **hbase.coprocessor.region.classes**, **hbase.coprocessor.master.classes**, and **hbase.coprocessor.regionserver.classes**.

Add the coprocessor parameter **org.apache.hadoop.hbase.security.access.AccessController** to the end of the values of the preceding parameters, and use a comma (,) to separate the values from those of the original coprocessors.

Step 5 Click **Save**, click **OK**, and wait for message "Operation successful" to display.

----End

Disabling Permission Verification on HBase

NOTE

After HBase permission verification is disabled, the existing permission data will be retained. If you want to delete permission information, disable permission verification, enter the HBase shell, and delete table **hbase:acl**.

Step 1 Log in to FusionInsight Manager.

Step 2 Click **Cluster**, click the name of the desired cluster, choose **Services** > **HBase**, and click **Configurations**.

Step 3 Click **All Configurations**.

Step 4 Search for parameters **hbase.coprocessor.region.classes**, **hbase.coprocessor.master.classes**, and **hbase.coprocessor.regionserver.classes**.

Delete the coprocessor parameter **org.apache.hadoop.hbase.security.access.AccessController**.

Step 5 Click **Save**, click **OK**, and wait for message "Operation successful" to display.

----End

Disabling Permission Verification on HDFS

Step 1 Log in to FusionInsight Manager.

Step 2 Click **Cluster**, click the name of the desired cluster, choose **Services** > **HDFS**, and click **Configurations**.

Step 3 Click **All Configurations**.

Step 4 Search for parameters **dfs.namenode.acls.enabled** and **dfs.permissions.enabled**.

- **dfs.namenode.acls.enabled** indicates whether to enable HDFS ACL. The default value is **true**, indicating that the ACL is enabled. Change the value to **false**.
- **dfs.permissions.enabled** indicates whether to enable permission check for HDFS. The default value is **true**, indicating that permission check is enabled. Change the value to **false**. After the modification, the owner, owner group, and permission of the directories and files in HDFS remain unchanged.

Step 5 Click **Save**, click **OK**, and wait for message "Operation successful" to display.

----End

Enabling Permission Verification on YARN

Step 1 Log in to FusionInsight Manager.

Step 2 Click **Cluster**, click the name of the desired cluster, choose **Services** > **Yarn**, and click **Configurations**.

Step 3 Click **All Configurations**.

Step 4 Search for parameter **yarn.acl.enable**.

yarn.acl.enable indicates whether to enable the permission check for YARN.

- In normal clusters, the value is set to **false** by default to disable permission check. To enable permission check, change the value to **true**.
- In security clusters, the value is set to **true** by default to enable authentication.

Step 5 Click **Save**, click **OK**, and wait for message "Operation successful" to display.

----End

Disabling Permission Verification on ZooKeeper

Step 1 Log in to FusionInsight Manager.

Step 2 Click **Cluster**, click the name of the desired cluster, choose **Services** > **ZooKeeper**, and click **Configurations**.

Step 3 Click **All Configurations**.

Step 4 Search for parameter **skipACL**.

skipACL indicates whether to skip the ZooKeeper permission check. The default value is **no**, indicating that permission check is enabled. Change the value to **yes**.

Step 5 Click **Save**, click **OK**, and wait for message "Operation successful" to display.

----End

9.12.2.1.4 Logging In to a Non-Cluster Node Using a Cluster User in Normal Mode

Scenario

When the cluster is installed in normal mode, the component clients do not support security authentication and cannot use the **kinit** command. Therefore, nodes outside the cluster cannot use users in the cluster by default. This may result in a user authentication failure when one of these nodes access a component server.

The node administrator can configure a user who has the same name as that of a user for a node outside the cluster, allow the user to log in to the node using the SSH protocol, and connect to the servers of components in the cluster by using the user who logs in to the OS.

Prerequisites

- Nodes outside the cluster can connect to the service plane of the cluster.
- The KrbServer service of the cluster is running properly.
- You have obtained the password of user **root** of the node outside the cluster.
- A human-machine user has been planned and added to the cluster, and you have obtained the authentication credential file. For details, see **Creating a User** and **Exporting an Authentication Credential File**.

Procedure

Step 1 Log in to the node where a user is to be added as user **root**.

Step 2 Run the following command:

rpm -qa | grep pam and **rpm -qa| grep krb5-client**

The following RPM packages are displayed:

```
pam_krb5-32bit-2.3.1-47.12.1
pam-modules-32bit-11-1.22.1
yast2-pam-2.17.3-0.5.211
pam-32bit-1.1.5-0.10.17
pam_mount-32bit-0.47-13.16.1
pam-config-0.79-2.5.58
pam_krb5-2.3.1-47.12.1
pam-doc-1.1.5-0.10.17
pam-modules-11-1.22.1
pam_mount-0.47-13.16.1
pam_ldap-184-147.20
pam-1.1.5-0.10.17
krb5-client-1.6.3
```

Step 3 Check whether the RPM packages in the list are installed in the OS.

- If yes, go to **Step 5**.
- If no, go to **Step 4**.

Step 4 Obtain the lacked RPM packages from the OS image, upload the files to the current directory, and run the following command to install the RPM package:

rpm -ivh *.rpm

NOTE

The RPM packages to be installed may bring security risks. The risks that may be brought by the installation of these RPM packages must be taken into consideration during OS hardening.

After the RPM packages are installed, go to **Step 5**.

Step 5 Run the following command to configure Kerberos authentication on PAM:

pam-config --add --krb5

NOTE

If you need to cancel Kerberos authentication and system user login on a non-cluster node, run the **pam-config --delete --krb5** command as user **root**.

Step 6 Decompress the authentication credential file to obtain **krb5.conf**, use WinSCP to upload this configuration file to the **/etc** directory on the node outside the cluster, and run the following command to configure related permission to enable other users to access the file, such as permission **604**:

chmod 604 /etc/krb5.conf

Step 7 Run the following command in the connection session as user **root** to add the corresponding OS user to the human-machine user, and specify **root** as the primary group.

The OS user password is the same as the initial password when the human-machine user is created on Manager.

```
useradd User name -m -d /home/admin_test -g root -s /bin/bash
```

For example, if the name of the human-machine user is **admin_test**, run the following command:

```
useradd admin_test -m -d /home/admin_test -g root -s /bin/bash
```

NOTE

When you use the newly added OS user to log in to the node by using the SSH protocol for the first time, the system prompts that the password has expired after you enter the user password, and the system prompts that the password needs to be changed after you enter the user password again. You need to enter a new password that meets the password complexity requirements of both the node OS and the cluster.

----End

9.12.2.2 Changing the Password for a System User

9.12.2.2.1 Changing the Password for User admin

Scenario

User **admin** is the system administrator account of FusionInsight Manager. You are advised to periodically change the password on FusionInsight Manager to improve system security.

Procedure

Step 1 Log in to FusionInsight Manager.

User **admin** is required for login.

Step 2 Move the cursor to **Hello, admin** in the upper right corner of the page.

In the displayed menu, click **Change Password**.

Step 3 Set **Old Password**, **New Password**, and **Confirm Password**, and click **OK**.

The password must meet the following complexity requirements:

- Contains 8 to 64 characters.
- Contains at least four types of the following: uppercase letters, lowercase letters, numbers, spaces, and special characters (~`!?,.;-_'(){}[]/<>@#$%^&*+|\=).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be a common easily-cracked password.
- Cannot be the same as the password used in the last *N* times. *N* indicates the value of **Repetition Rule** in Configuring Password Policies.

----End

9.12.2.2.2 Changing the Password for an OS User

Scenario

During FusionInsight Manager installation, the system automatically creates user **omm** and **ommdba** on each node in the cluster. Periodically change the login passwords of the OS users **omm** and **ommdba** of the cluster node to improve the system O&M security.

The passwords of users **omm** and **ommdba** of the nodes can be different.

Prerequisites

- You have obtained the IP address of the node where the passwords of users **omm** and **ommdba** are to be changed.
- You have obtained the password of user **root** before changing the passwords of users **omm** and **ommdba**.

Changing the Password of an OS User

Step 1 Log in to the node where the password is to be changed as user **root**.

Step 2 Run the following command to change the user password:

passwd *ommdba*

Red Hat system displays the following information:

```
Changing password for user ommdba.
New password:
```

Step 3 Enter a new password. The policy for changing the password of an OS user varies according to the OS that is actually used.

```
Retype New Password:
Password changed.
```

----End

9.12.2.3 Changing the Password for a System Internal User

9.12.2.3.1 Changing the Password for the Kerberos Administrator

Scenario

It is recommended that the administrator periodically change the password of Kerberos administrator **kadmin** to improve the system O&M security.

If the user password is changed, the OMS Kerberos administrator password is changed as well.

Prerequisites

You have installed the client on any node in the cluster and obtained the IP address of the node.

Procedure

Step 1 Log in to the node where the client is installed as user **root**.

Step 2 Run the following command to go to the client directory, for example, **/opt/hadoopclient**:

```
cd /opt/hadoopclient
```

Step 3 Run the following command to set environment variables:

```
source bigdata_env
```

Step 4 Run the following command to change the password for **kadmin/admin**. The password changing takes effect on all servers.

```
kpasswd kadmin/admin
```

The password must meet the following complexity requirements:

- Contains at least 8 characters.
- Contains at least four types of the following: uppercase letters, lowercase letters, numbers, spaces, and special characters (~`!?,.;-_'(){}[]/<>@#$%^&*+|\=).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be a common easily-cracked password, for example, **Admin@12345**.
- Cannot be the same as the password used in the last *N* times. *N* indicates the value of **Repetition Rule** in **Configuring Password Policies**.

----End

9.12.2.3.2 Changing the Password for the OMS Kerberos Administrator

Scenario

It is recommended that the administrator periodically change the password of OMS Kerberos administrator **kadmin** to improve the system O&M security.

If the user password is changed, the Kerberos administrator password is changed as well.

Procedure

Step 1 Log in to any management node in the cluster as user **omm**.

Step 2 Run the following command to go to the related directory:

```
cd ${BIGDATA_HOME}/om-server/om/meta-0.0.1-SNAPSHOT/kerberos/scripts
```

Step 3 Run the following command to set environment variables:

```
source component_env
```

Step 4 Run the following command to change the password for **kadmin/admin**. This operation takes effect for all servers.

```
kpasswd kadmin/admin
```

The password must meet the following complexity requirements:

- Contains at least 8 characters.
- Contains at least four types of the following: uppercase letters, lowercase letters, numbers, and special characters (~`!?,.;-_'(){}[]/<>@#$%^&*+|\=).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be a common easily-cracked password, for example, **Admin@12345**.
- Cannot be the same as the password used in the last *N* times. *N* indicates the value of **Repetition Rule** in **Configuring Password Policies**.

----End

9.12.2.3.3 Changing the Passwords of the LDAP Administrator and the LDAP User (Including OMS LDAP)

NOTE

This section applies only to MRS 3.1.0. For later versions, see **Modifying OMS Service Configuration Parameters**.

Scenario

It is recommended that the administrator periodically changes the passwords of LDAP administrator **cn=root,dc=hadoop,dc=com** and LDAP user **cn=pg_search_dn,ou=Users,dc=hadoop,dc=com** to improve the system O&M security.

If the passwords are changed, the password of the OMS LDAP administrator or user is changed as well.

NOTE

If the cluster is upgraded from an early version to a latest version, the LDAP administrator password will inherit the password policy of the old cluster. To ensure system security, you are advised to change the password after the cluster upgrade.

Impact on the System

- Changing the user password of the LdapServer service is a high-risk operation and requires restarting the KrbServer and LdapServer services. If KrbServer is restarted, users may fail to be queried by running the **id** command on nodes in the cluster temporarily. Therefore, exercise caution when restarting KrbServer.
- After the password of LDAP user **cn=pg_search_dn,ou=Users,dc=hadoop,dc=com** is changed, the user may be locked in the LDAP component. Therefore, you are advised to unlock the user after changing the password. For details about how to unlock the user, see **Unlocking LDAP Users and Management Accounts**.

Prerequisites

Before changing the password of LDAP user **cn=pg_search_dn,ou=Users,dc=hadoop,dc=com**, ensure that the user is not locked by running the following command on the active management node of the cluster:

> **NOTE**
>
> To query the OLdap port number, perform the following steps:
>
> 1. Log in to FusionInsight Manager, choose **System** > **OMS** > **oldap** > **Modify Configuration**:
> 2. The value of **LDAP Service Listening Port** is the OLDAP port.

ldapsearch -H ldaps://*Floating IP address of OMS:OLDAP port***-LLL -x -D cn=pg_search_dn,ou=Users,dc=hadoop,dc=com -W -b cn=pg_search_dn,ou=Users,dc=hadoop,dc=com -e ppolicy**

Enter the password of the LDAP user **pg_search_dn**. If the following information is displayed, the user is locked. In this case, unlock the user. For details, see **Unlocking LDAP Users and Management Accounts**.

> **NOTE**
>
> The password of the LDAP user **pg_search_dn** is randomly generated by the system. You can obtain the password from the **/etc/sssd/sssd.conf or /etc/ldap.conf** file on the active node.

```
ldap_bind: Invalid credentials (49); Account locked
```

Procedure

Step 1 Log in to FusionInsight Manager, click **Cluster**, click the name of the desired cluster, and choose **Service** > **LdapServer**.

Step 2 Choose **More** > **Change Database Password**. In the displayed dialog box, enter the password of the current login user and click **OK**.

Step 3 In the **Change Password** dialog box, select the user whose password to be modified in the **User Information** drop-down box.

Step 4 Enter the old password in the **Old Password** text box, and enter the new password in the **New Password** and **Confirm Password** text boxes.

The password must meet the following complexity requirements:

- Contains 16 to 32 characters.
- Contains at least three types of the following: uppercase letters, lowercase letters, numbers, spaces, and special characters (`~!@#$%^&*()-_=+|[{}];,<.>/?).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be the same as the current password.

Step 5 Select **I have read the information and understood the impact** and click **OK** to confirm the modification and restart the service.

----End

9.12.2.3.4 Changing the Password for the LDAP Administrator

NOTE

This section applies only to MRS 3.1.0. For later versions, see **Modifying OMS Service Configuration Parameters**.

Scenario

It is recommended that the administrator periodically changes the passwords of LDAP administrator accounts **cn=krbkdc,ou=Users,dc=hadoop,dc=com** and **cn=krbadmin,ou=Users,dc=hadoop,dc=com** to improve the system O&M security.

Impact on the System

- You need to restart the KrbServer service after changing the password.
- After the password is changed, check whether the LDAP administrator accounts **cn=krbkdc,ou=Users,dc=hadoop,dc=com** and **cn=krbadmin,ou=Users,dc=hadoop,dc=com** are locked, run the following command on the active management node of the cluster to check whether **krbkdc** is locked (the method for user **krbadmin** is similar):

  NOTE

  OLdap port number obtaining method:

  1. Log in to FusionInsight Manager, choose **System** > **OMS** > **oldap** > **Modify Configuration**:
  2. The **LDAP Listening Port** parameter value is **oldap port**.

  ldapsearch -H ldaps://*OMS_FLOAT_ IP address:OLdap port* **-LLL -x -D cn=krbkdc,ou=Users,dc=hadoop,dc=com -W -b cn=krbkdc,ou=Users,dc=hadoop,dc=com -e ppolicy**

  Enter the password of the LDAP administrator account **krbkdc**. If the following message is displayed, the account is locked. For details about how to unlock the account, see **Unlocking LDAP Users and Management Accounts**.

  ```
  ldap_bind: Invalid credentials (49); Account locked
  ```

Prerequisites

You have obtained the management node IP address.

Procedure

Step 1 Log in to the active management node as user **omm** with the IP address of the active management node.

Step 2 Run the following command to go to the related directory:

cd ${BIGDATA_HOME}/om-server/om/meta-0.0.1-SNAPSHOT/kerberos/scripts

Step 3 Run the following command to change the password of the LDAP administrator account:

./okerberos_modpwd.sh

Enter the old password and then enter a new password twice.

The password must meet the following complexity requirements:

- Contains 16 to 32 characters.
- Contains at least three types of the following: uppercase letters, lowercase letters, numbers, spaces, and special characters (`~!@#$%^&*()-_=+|[{}];,<.>/?).
- Cannot be the same as the current password.

If the following information is displayed, the password is changed.

```
Modify kerberos server password successfully.
```

Step 4 Log in to FusionInsight Manager, click **Cluster**, click the name of the desired cluster, and choose **Services** > **KrbServer**. On the displayed page, choose **More** > **Restart Service**.

Enter the password and do not select **Restart upper-layer services**. Click **OK** to restart the KrbServer service.

----End

9.12.2.3.5 Changing the Password for a Component Running User

Scenario

It is recommended that the administrator periodically change the password for each component running user to improve the system O&M security.

Component running users can be classified into the following two types depending on whether their initial passwords are randomly generated by the system:

- If the initial password of a component running user is randomly generated by the system, the user is of the machine-machine type.
- If the initial password of a component running user is not randomly generated by the system, the user is of the human-machine type.

Impact on the System

If the initial password is randomly generated by the system, the cluster needs to be restarted for the password changing to take effect. Services are unavailable during the restart.

Prerequisites

You have installed the client on any node in the cluster and obtained the IP address of the node.

Procedure

Step 1 Log in to the node where the client is installed as the client installation user

Step 2 Run the following command to switch to the client directory, for example, **/opt/client**:

cd /opt/client

Step 3 Run the following command to set environment variables:

```
source bigdata_env
```

Step 4 Run the following command and enter the password of user **kadmin/admin** to log in to the **kadmin** console:

```
kadmin -p kadmin/admin
```

NOTE

The default password of user **kadmin/admin**, **Admin@123**, will expire upon your first login. Change the password as prompted and keep the new password secure.

Step 5 Run the following command to change the password of an internal component running user. The password changing takes effect on all servers.

```
cpw Internal system username
```

For example: **cpw oms/manager**

The password must meet the following complexity requirements:

- Contains at least 8 characters.
- Contains at least four types of the following: uppercase letters, lowercase letters, numbers, spaces, and special characters (~`!?,.;-_'(){}[]/<>@#$%^&*+|\=).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be a common easily-cracked password, for example, **Admin@12345**.
- Cannot be the same as the password used in latest *N* times. *N* indicates the value of **Number of Historical Passwords** configured in **Configuring Password Policies**. This policy applies to only human-machine accounts.

NOTE

Run the following command to check user information:

```
getprinc Internal system username
```

For example: **getprinc oms/manager**

Step 6 Determine the type of the user whose password needs to be changed.

- If the user is a machine-machine user, go to **Step 7**.
- If the user is a human-machine user, the password is changed successfully and no further action is required.

Step 7 Log in to FusionInsight Manager.

Step 8 Click **Cluster**, click the name of the desired cluster, and choose **More** > **Restart**.

Step 9 In the displayed window, enter the password of the current login user and click **OK**.

Step 10 In the displayed restart confirmation dialog box, click **OK**.

Step 11 Wait for message "Operation successful" to display.

----End

9.12.2.4 Changing the Password for a Database User

9.12.2.4.1 Changing the Password of the OMS Database Administrator

Scenario

It is recommended that the administrator periodically change the password of the OMS database administrator to improve the system O&M security.

Procedure

Step 1 Log in to the active management node as user **root**.

NOTE

The password of user **ommdba** cannot be changed on the standby management node. Otherwise, the cluster may not work properly. Change the password on the active management node only.

Step 2 Run the following command to switch to another user:

su - omm

Step 3 Run the following command to go to the related directory:

cd $OMS_RUN_PATH/tools

Step 4 Run the following command to change the password for user **ommdba**:

mod_db_passwd ommdba

Step 5 Enter the old password of user **ommdba** and enter a new password twice.

The password must meet the following complexity requirements:

- Contains 16 to 32 characters.
- Contains at least three types of the following: uppercase letters, lowercase letters, numbers, and special characters (~`!@#$%^&*()-+_=\|[{}];",<.>/?).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be the same as the last 20 historical passwords.

If the following information is displayed, the password is changed.

```
Congratulations, update [ommdba] password successfully.
```

----End

9.12.2.4.2 Changing the Password for the Data Access User of the OMS Database

Scenario

It is recommended that the administrator periodically change the password of the user accessing the OMS database to improve the system O&M security.

Impact on the System

The OMS service needs to be restarted for the new password to take effect. The service is unavailable during the restart.

Procedure

Step 1 On FusionInsight Manager, choose **System** > **OMS** > **gaussDB** > **Change Password**.

Step 2 Locate the row where user **omm** is located and click **Change Password** in the **Operation** column.

Step 3 In the displayed window, enter the password of the current login user and click **OK**.

Step 4 Enter the old and new passwords as prompted.

The password must meet the following complexity requirements:

- Contains 8 to 32 characters.
- Contains at least three types of the following: uppercase letters, lowercase letters, numbers, and special characters (~`!@#$%^&*()-+_=\|[{}];",<.>/?).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be the same as the last 20 historical passwords.

Step 5 Click **OK**. Wait until the system displays a message indicating that the operation is successful.

Step 6 Locate the row where user **omm** is located and click **Restart OMS Service** in the **Operation** column.

Step 7 In the displayed window, enter the password of the current login user and click **OK**.

Step 8 In the displayed restart confirmation dialog box, click **OK** to restart the OMS service.

----End

9.12.2.4.3 Changing the Password for a Component Database User

Scenario

It is recommended that the administrator periodically change the password for each component database user to improve the system O&M security.

NOTE

This section applies only to MRS 3.1.0. For versions later than MRS 3.1.0, see **Resetting the Component Database User Password**.

Impact on the System

The services need to be restarted for the new password to take effect. The services are unavailable during the restart.

Procedure

Step 1 On FusionInsight Manager, click **Cluster**, click the name of the desired cluster, and click **Services**.

Step 2 Determine the component database user whose password is to be changed.

For details about how to change the password of database user **omm** of DBService, perform operations in **Changing the Password for User omm in DBService**. To change the passwords of database users of other components, you need to stop services first and then perform the operations in **Step 3**.

Step 3 Click the service whose database user password is to be changed, and choose **More** > **Change Database Password**. On the displayed page, enter the password of the current login user and click **OK**.

Step 4 Enter the old and new passwords as prompted.

The password must meet the following complexity requirements:

- Contains 8 to 32 characters.
- Contains at least three types of the following: uppercase letters, lowercase letters, numbers, and special characters (~`!@#$%^&*()-+_=\|[{}];",<.>/?).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be the same as the last 20 historical passwords.

Step 5 Select **I have read the information and understand the impact** and click **OK**.

Step 6 After the password is changed, choose **More** > **Restart Service**. In the displayed dialog box, enter the password of the current login user, click **OK**, and select **Restart the upper-layer services**. Click **OK** to restart the services.

----End

9.12.2.4.4 Resetting the Component Database User Password

Scenario

Default passwords for components in the MRS cluster to connect to the DBService database are random. You are advised to periodically reset the passwords of component database users to improve system O&M security.

NOTE

This section applies only to MRS 3.1.2 or later. For versions earlier than MRS 3.1.2, see **Changing the Password for a Component Database User**.

Impact on the System

To reset passwords, you need to stop and then restart services, during which services are unavailable.

Procedure

Step 1 On FusionInsight Manager, click **Cluster**, click the name of the desired cluster, and click **Services**.

Step 2 Click the name of the service whose database user password is to be reset, for example, **Kafka**, and click **Stop Service** on the **Dashboard** page.

In the displayed dialog box, enter the password of the current login user and click **OK**.

After confirming the impact of stopping the service, wait until the service is stopped.

Step 3 On the **Dashboard** page, choose **More** > **Reset Database Password**.

In the displayed dialog box, enter the password of the current login user and click **OK**.

Select "I have read the information and understand the impact", and click **OK**.

Step 4 After the password is reset, click **Start Service** on the **Dashboard** page.

Step 5 In the displayed dialog box, click **OK** and wait until the service is started.

----End

9.12.2.4.5 Changing the Password for User omm in DBService

Step 1 Log in to the active DBService node as user **root**.

NOTE

The password of user **omm** for the DBService database cannot be changed on the standby DBService node. Change the password on the active DBService node only.

Step 2 Run the following command to switch to another user:

su - omm

Step 3 Run the following command to go to the related directory:

source $DBSERVER_HOME/.dbservice_profile

cd ${DBSERVICE_SOFTWARE_DIR}/sbin/

Step 4 Run the following command to change the password of user **omm**:

sh modifyDBPwd.sh

Step 5 Enter the old password of user **omm** and enter a new password twice.

The password must meet the following complexity requirements:

- Contains 8 to 32 characters.
- Contains at least three types of the following: uppercase letters, lowercase letters, numbers, and special characters (~`!@#$%^&*()-+_=\|[{}];",<.>/?).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be the same as the last 20 historical passwords.

If the following information is displayed, the password is changed.

```
Successful to modify password.
```

----End

9.12.2.4.6 Changing the Password for User compdbuser of the DBService Database

Scenario

It is recommended that the administrator periodically change the password of the OMS database administrator to improve the system O&M security.

Procedure

Step 1 Log in to FusionInsight Manager, choose **Cluster** > **Services** > **DBService**, click **Instance**, and view the IP address of the active DBService node.

Step 2 Log in to the active DBService node as user **root**.

> **NOTE**
>
> The password of user **compuserdb** cannot be changed on the standby DBService node. Change the password on the active management node only.

Step 3 Switch to the **$DBSERVER_HOME** directory and configure environment variables:

su - omm

cd $DBSERVER_HOME

source .dbservice_profile

Step 4 Run the following command to change the password of user **compdbuser** as user **omm** of the DBService database:

gsql -U omm -W omm*Password of user omm of the DBService database* **-d postgres -p 20051 -c "alter user compdbuser identified by '***New password'* **valid until '***Expiration time***';"**

> **NOTE**
>
> - The new password must meet the following complexity requirements:
>   - Contains 16 to 32 characters.
>   - Contains at least three types of the following: uppercase letters, lowercase letters, numbers, and special characters (~`!@#$%^&*()-+_=\|[{}];:",<.>/?).
>   - Cannot be the same as the username or the username spelled backwards.
>   - Cannot be the same as the last 20 historical passwords.
> - The expiration time format is xxxx-xx-xx, for example, **2020-10-31**.

If the following information is displayed, the modification is successful:

```
ALTER ROLE
```

----End

9.12.3 Security Hardening

9.12.3.1 Hardening Policies

Hardening Tomcat

Tomcat is hardened as follows based on open-source software during FusionInsight Manager software installation and use:

- The Tomcat version is upgraded to the official version.
- Permissions on the directories under applications are set to **500**, and the write permission on some directories is supported.
- The Tomcat installation package is automatically deleted after the system software is installed.
- The automatic deployment function is disabled for projects in application directories. Only the **web**, **cas**, and **client** projects are deployed.
- Some unused **http** methods are disabled, preventing attacks by using the **http** methods.
- The default shutdown port and command of the Tomcat server are changed to prevent hackers from shutting down the server and attacking servers and applications.
- To ensure security, the value of **maxHttpHeaderSize** is changed, which enables server administrators to control abnormal requests of clients.
- The Tomcat version description file is modified after Tomcat is installed.
- To prevent disclosure of Tomcat information, the Server attributes of Connector are modified so that attackers cannot obtain information about the server.
- Permissions on files and directories of Tomcat, such as the configuration files, executable files, log directories, and temporary folders, are under control.
- Session facade recycling is disabled to prevent request leakage.
- LegacyCookieProcessor is used as CookieProcessor to prevent the leakage of sensitive data in cookies.

Hardening LDAP

LDAP is hardened as follows after a cluster is installed:

- In the LDAP configuration file, the password of the administrator account is encrypted using SHA. After the OpenLDAP is upgraded to 2.4.39 or later, data is automatically synchronized between the active and standby LDAP nodes using the SASL External mechanism, which prevents disclosure of the password.
- The LDAP service in the cluster supports the SSLv3 protocol by default, which can be used safely. When the OpenLDAP is upgraded to 2.4.39 or later, the LDAP automatically uses TLS1.0 or later to prevent unknown security risks.

Hardening JDK

- If the client process uses the AES256 encryption algorithm, JDK security hardening is required. The operations are as follows:

  Obtain the Java Cryptography Extension (JCE) package whose version matches that of JDK. The JCE package contains **local_policy.jar** and

US_export_policy.jar. Copy the JAR files to the following directory and replace the files in the directory.

- Linux: *JDK installation directory***/jre/lib/security**
- Windows: *JDK installation directory***\jre\lib\security**

NOTE

Access the Open JDK open-source community to obtain the JCE file.

- If the client process uses the SM4 encryption algorithm, the JAR package needs to be updated.

  Obtain **SMS4JA.jar** in the *client installation directory***/JDK/jdk/jre/lib/ext/** directory, and copy the JAR package to the following directory:

  - Linux: *JDK installation directory***/jre/lib/ext/**
  - Windows: *JDK installation directory***\jre\lib\ext\**

9.12.3.2 Configuring a Trusted IP Address to Access LDAP

Scenario

By default, the LDAP service deployed in the OMS and cluster can be accessed by any IP address. To enable the LDAP service to be accessed by only trusted IP addresses, you can configure the INPUT policy in the iptables filtering list.

Impact on the System

After the configuration, the LDAP service cannot be accessed by IP addresses that are not configured. Before the expansion, the added IP addresses need to be configured as trusted IP addresses.

Prerequisites

- You have collected the management plane IP addresses and service plane IP addresses of all nodes in the cluster and all floating IP addresses.
- You have obtained the **root** user account for all nodes in the cluster.

Procedure

Configuring trusted IP addresses for the LDAP service on the OMS

Step 1 Confirm the management node IP address. For details, see **Logging In to the Management Node**.

Step 2 Log in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Step 3 Choose **System** > **OMS** and choose **oldap** > **Modify Configuration** to view the OMS LDAP port number, that is, the value of **LDAP Listening Port**. The default port number is **21750**.

Step 4 Log in to the active management node as user **root** using the IP address of the active management node.

Step 5 Run the following command to check the INPUT policy in the iptables filtering list:

iptables -L

For example, if no rule is configured, the INPUT policy is displayed as follows:

```
Chain INPUT (policy ACCEPT)
target     prot opt source               destination
```

Step 6 Run the following command to configure all IP addresses used by the cluster as trusted IP addresses. Each IP address needs to be added independently.

iptables -A INPUT -s *Trusted IP address* **-p tcp --dport** *Port number* **-j ACCEPT**

For example, to configure **10.0.0.1** as a trusted IP address and enable it to access port **21750**, you need to run the following command:

iptables -A INPUT -s 10.0.0.1 -p tcp --dport 21750 -j ACCEPT

Step 7 Run the following command to configure all IP addresses as untrusted IP addresses. The trusted IP addresses will not be affected by this rule.

iptables -A INPUT -p tcp --dport *Port number* **-j DROP**

For example, to disable all IP addresses to access port **21750**, run the following command:

iptables -A INPUT -p tcp --dport 21750 -j DROP

Step 8 Run the following command to view the modified INPUT policy in the iptables filtering list:

iptables -L

For example, after a trusted IP address is configured, the INPUT policy is displayed as follows:

```
Chain INPUT (policy ACCEPT)
target     prot opt source               destination
ACCEPT     tcp  --  10.0.0.1             anywhere            tcp dpt:21750
DROP       tcp  --  anywhere             anywhere            tcp dpt:21750
```

Step 9 Run the following command to view the rules and rule numbers in the iptables filtering list:

iptables -L -n --line-number

```
Chain INPUT (policy ACCEPT)
num target     prot opt source               destination
1   DROP       tcp  --  0.0.0.0/0            0.0.0.0/0           tcp dpt:21750
```

Step 10 Run the following command to delete the desired rule from the iptables filtering list based on site requirement:

iptables -D INPUT *Number of the rule to be deleted*

For example, to delete rule 1, run the following command:

iptables -D INPUT 1

Step 11 Log in to the standby management node as user **root** using the standby IP address. Repeat **Step 5** to **Step 10**.

Configuring trusted IP addresses for the LDAP service in the cluster

Step 12 Log in to FusionInsight Manager.

Step 13 Click **Cluster**, click the name of the desired cluster, and choose **Service** > **LdapServer**. On the displayed page, click **Instance** to view the nodes where the LDAP services locate.

Step 14 Go to the **Configurations** page, and view the LDAP port number of the cluster, that is, the value of **LDAP_SERVER_PORT**. The default value is **21780**.

Step 15 Log in to the LDAP node as user **root** using the LDAP service IP address.

Step 16 Run the following command to view the INPUT policy in the iptables filtering list:

iptables -L

For example, if no rule is configured, the INPUT policy is displayed as follows:

```
Chain INPUT (policy ACCEPT)
target     prot opt source               destination
```

Step 17 Run the following command to configure all IP addresses used by the cluster as trusted IP addresses. Each IP address needs to be added independently.

iptables -A INPUT -s *Trusted IP address* **-p tcp --dport** *Port number* **-j ACCEPT**

For example, to configure **10.0.0.1** as a trusted IP address and enable it to access port **21780**, you need to run the following command:

iptables -A INPUT -s 10.0.0.1 -p tcp --dport 21780 -j ACCEPT

Step 18 Run the following command to configure all IP addresses as untrusted IP addresses. The trusted IP addresses will not be affected by this rule.

iptables -A INPUT -p tcp --dport *Port number* **-j DROP**

For example, to disable all IP addresses to access port **21780**, run the following command:

iptables -A INPUT -p tcp --dport 21780 -j DROP

Step 19 Run the following command to view the modified INPUT policy in the iptables filtering list:

iptables -L

For example, after a trusted IP address is configured, the INPUT policy is displayed as follows:

```
Chain INPUT (policy ACCEPT)
target     prot opt source               destination
ACCEPT     tcp  --  10.0.0.1             anywhere            tcp dpt:21780
DROP       tcp  --  anywhere             anywhere            tcp dpt:21780
```

Step 20 Run the following command to view the rules and rule numbers in the iptables filtering list:

iptables -L -n --line-number

```
Chain INPUT (policy ACCEPT)
num target     prot opt source               destination
1   DROP       tcp  --  0.0.0.0/0            0.0.0.0/0           tcp dpt:21780
```

Step 21 Run the following command to delete the desired rule from the iptables filtering list based on site requirement:

iptables -D INPUT *Number of the rule to be deleted*

For example, to delete rule 1, run the following command:

iptables -D INPUT 1

Step 22 Log in to the LDAP node as user **root** using the IP address of another LDAP service, and repeat **Step 16** to **Step 21**.

----End

9.12.3.3 HFile and WAL Encryption

HFile and WAL Encryption

NOTICE

- Setting the HFile and WAL encryption mode to SMS4 or AES has a great impact on the system and will cause data loss in case of any misoperation. Therefore, this operation is not recommended.
- Batch data import using Bulkload does not support data encryption.

HFile and Write ahead log (WAL) in HBase are not encrypted by default. To encrypt them, perform the following operations.

Step 1 On any HBase node, run the following commands to create a key file as user **omm**:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.1.0.1/install/FusionInsight-HBase-2.2.3/hbase/bin/hbase-encrypt.sh ***<path>/hbase.jks <type> <length> <alias>***

- */<path>/hbase.jks* indicates the path for storing the generated JKS file.
- *<type>* indicates the encryption type, which can be SMS4 or AES.
- *<length>* indicates the key length. SMS4 supports 16-bit and AES supports 128-bit.
- *<alias>* indicate the alias of the key file. When you create the key file for the first time, retain the default value **omm**.

For example, to generate an SMS4 encryption key, run the following command:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.1.0.1/install/FusionInsight-HBase-2.2.3/hbase/bin/hbase-encrypt.sh /home/hbase/conf/hbase.jks SMS4 16 omm

To generate an AES encryption key, run the following command:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.1.0.1/install/FusionInsight-HBase-2.2.3/hbase/bin/hbase-encrypt.sh /home/hbase/conf/hbase.jks AES 128 omm

NOTE

- To ensure operations can be successfully performed, the **<path>/hbase.jks** directory needs to be created in advance, and the cluster operation user must have the **rw** permission of this directory.
- After running the command, enter the same *<password>* four times. The password encrypted in **Step 3** is the same as the password in this step.

Step 2 Distribute the generated key files to the same directory on all nodes in the cluster and assign read and write permission to user **omm**.

NOTE

- Administrators need to select a safe procedure to distribute keys based on the enterprise security requirements.
- If the key files of some nodes are lost, repeat the step to copy the key files from other nodes.

Step 3 On FusionInsight Manager, set **hbase.crypto.keyprovider.parameters.encryptedtext** to the encrypted password. Set **hbase.crypto.keyprovider.parameters.uri** to the path and name of the key file.

- The format of **hbase.crypto.keyprovider.parameters.uri** is **jceks://** *<key_Path_Name>*.

  <key_Path_Name> indicates the path of the key file. For example, if the path of the key file is **/home/hbase/conf/hbase.jks**, set this parameter to **jceks:///home/hbase/conf/hbase.jks**.
- The format of **hbase.crypto.keyprovider.parameters.encryptedtext** is *<encrypted_password>*.

  <encrypted_password> indicates the encrypted password generated during the key file creation. The parameter value is displayed in ciphertext. Run the following command as user **omm** to obtain the related encrypted password on the nodes where HBase service is installed:

  sh ${BIGDATA_HOME}/FusionInsight_HD_8.1.0.1/install/FusionInsight-HBase-2.2.3/hbase/bin/hbase-encrypt.sh

  NOTE

  After running the command, you need to enter **<password>**. The password is the same as that entered in **Step 1**.

Step 4 On FusionInsight Manager, set **hbase.crypto.key.algorithm** to **SMS4** or **AES** to use SMS4 or AES for HFile encryption.

Step 5 On FusionInsight Manager, set **hbase.crypto.wal.algorithm** to **SMS4** or **AES** to use SMS4 or AES for WAL encryption.

Step 6 On FusionInsight Manager, set **hbase.regionserver.wal.encryption** to **true**.

Step 7 Save the settings and restart the HBase service for the settings to take effect.

Step 8 Create an HBase table through CLI or code and configure the encryption mode to enable encryption. **<type>** indicates the encryption type, and **d** indicates the column family.

- When you create an HBase table through CLI, set the encryption mode to SMS4 or AES for the column family.

create '*<table name>*', {*NAME => 'd'*, ***ENCRYPTION =>*** '*<type>*'}

- When you create an HBase table using code, set the encryption mode to SMS4 or AES by adding the following information to the code:

```
public void testCreateTable()
{
    String tableName = "user";
    Configuration conf = getConfiguration();
    HTableDescriptor htd = new HTableDescriptor(TableName.valueOf(tableName));

    HColumnDescriptor hcd = new HColumnDescriptor("d");
    //Set the encryption mode to SMS4 or AES.
    hcd.setEncryptionType("<type>");
    htd.addFamily(hcd);

    HBaseAdmin admin = null;
    try
    {
        admin = new HBaseAdmin(conf);

        if(!admin.tableExists(tableName))
        {
            admin.createTable(htd);
        }
    }
    catch (IOException e)
    {
        e.printStackTrace();
    }
    finally
    {
        if(admin != null)
        {
            try
            {
                admin.close();
            }
            catch (IOException e)
            {
                e.printStackTrace();
            }
        }
    }
}
```

Step 9 If you have configured SMS4 or AES encryption by performing **Step 1** to **Step 7**, but do not set the related encryption parameter when creating the table in **Step 8**, the inserted data is not encrypted.

In this case, you can perform the following steps to encrypt the inserted data:

1. Run the **flush** command for the table to import the data in the memory to the HFile.

   flush '*<table_name>*'

2. Run the following commands to modify the table properties:

   disable '*<table_name>*'

   alter '*<table_name>*',**NAME=>**'*<column_name>*',**ENCRYPTION =>** '*<type>*'

   enable '*<table_name>*'

3. Insert a new data record and flush the table.

   NOTE

   A new data record must be inserted so that the HFile will generate a new HFile and the unencrypted data inserted previously will be rewritten and encrypted.

put *'<table_name>'*, **'id2','f1:c1','value22222222222222222222222222222222'**

flush *'<table_name>'*

4. Perform the following step to rewrite the HFile:

   major_compact'*<table_name>'*

NOTICE

During this step, the HBase table is disabled and cannot provide services. Exercise caution when you perform this step.

----End

Modifying a Key File

NOTICE

Modifying a key file has a great impact on the system and will cause data loss in case of any misoperation. Therefore, this operation is not recommended.

During the **HFile and WAL Encryption** operation, the related key file must be generated and its password must be set to ensure system security. After a period of running, you can replace the key file with a new one to encrypt HFile and WAL.

Step 1 Run the following command to generate a new key file as user **omm**:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.1.0.1/install/FusionInsight-HBase-2.2.3/hbase/bin/hbase-encrypt.sh *<path>/hbase.jks <type> <length> <alias-new>*

- *<path>/hbase.jks*: indicates the path for storing the generated **hbase.jks** file. The path and file name must be consistent with those of the key file generated in **HFile and WAL Encryption**.
- *<alias-new>*: indicates the alias of the key file. The alias must be different with that of the old key file.
- *<type>*: indicates the encryption type, which can be SMS4 or AES.
- *<length>* indicates the key length. SMS4 supports 16-bit and AES supports 128-bit.

For example, to generate an SMS4 encryption key, run the following command:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.1.0.1/install/FusionInsight-HBase-2.2.3/hbase/bin/hbase-encrypt.sh /home/hbase/conf/hbase.jks SMS4 16 omm_new

To generate an AES encryption key, run the following command:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.1.0.1/install/FusionInsight-HBase-2.2.3/hbase/bin/hbase-encrypt.sh /home/hbase/conf/hbase.jks AES 128 omm_new

NOTE

- To ensure operations can be successfully performed, the **<path>/hbase.jks** directory needs to be created in advance, and the cluster operation user must have the **rw** permission of this directory.
- After running the command, you need to enter the same *<password>* for three times. This password is the password of the key file. You can use the password of the old file without any security risk.

Step 2 Distribute the generated key files to the same directory on all nodes in the cluster and assign read and write permission to user **omm**.

NOTE

Administrators need to select a safe procedure to distribute keys based on the enterprise security requirements.

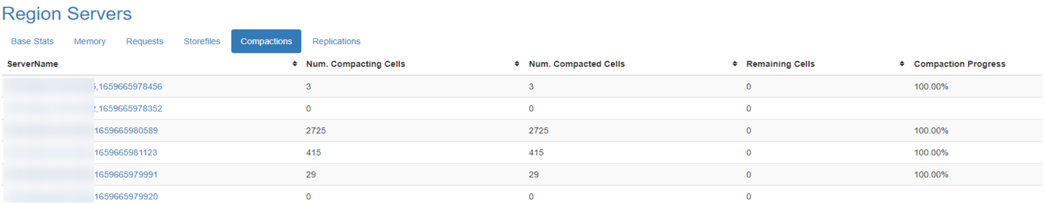
Step 3 On the HBase service configuration page of FusionInsight Manager, add custom configuration items, set **hbase.crypto.master.key.name** to **omm_new**, set **hbase.crypto.master.alternate.key.name** to **omm**, and save the settings.

Step 4 Restart the HBase service for the configuration to take effect.

Step 5 In HBase shell, run the **major compact** command to generate the HFile file based on the new encryption algorithm.

major_compact *'<table_name>'*

Step 6 You can view the major compact progress from the HMaster web page.



ServerName	Num. Compacting Cells	Num. Compacted Cells	Remaining Cells	Compaction Progress
[redacted],1659665978456	3	3	0	100.00%
[redacted],1659665978352	0	0	0	
[redacted]1659665980589	2725	2725	0	100.00%
[redacted]1659665981123	415	415	0	100.00%
[redacted]1659665979991	29	29	0	100.00%
[redacted]1659665979920	0	0	0	

Step 7 When all items in **Compaction Progress** reach **100%** and those in **Remaining KVs** are **0**, run the following command as user **omm** to destroy the old key file:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.1.0.1/install/FusionInsight-HBase-2.2.3/hbase/bin/hbase-encrypt.sh *<path>/hbase.jks <alias-old>*

- *<path>/hbase.jks*: indicates the path for storing the generated **hbase.jks** file. The path and file name must be consistent with those of the key file generated in **HFile and WAL Encryption**.
- *<alias-old>*: indicates the alias of the old key file to be deleted.

For example:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.1.0.1/install/FusionInsight-HBase-2.2.3/hbase/bin/hbase-encrypt.sh /home/hbase/conf/hbase.jks omm

NOTE

To ensure operations can be successfully performed, the **<path>/hbase.jks** directory needs to be created in advance, and the cluster operation user must have the **rw** permission of this directory.

Step 8 Repeat **Step 2** and distribute the updated key files again.

Step 9 Delete the HBase self-defined configuration item **hbase.crypto.master.alternate.key.name** added in **Step 3** from FusionInsight Manager.

Step 10 Repeat **Step 4** for the configuration take effect.

----End

9.12.3.4 Configuring Hadoop Security Parameters

Configuring Security Channel Encryption

The channels between components are not encrypted by default. You can set the following parameters to configure security channel encryption.

Page access for setting parameters: On FusionInsight Manager, click **Cluster**, click the name of the desired cluster, click **Services**, and click the target service. On the displayed page, click **Configuration** and click **All Configurations**. Enter a parameter name in the search box.

NOTE

Restart corresponding services for the modification to take effect after you modify configuration parameters.

Table 9-84 Parameter description

Service	Parameter	Description	Default Value
HBase	hbase.rpc.protection	Indicates whether the HBase channels, including the remote procedure call (RPC) channels for HBase clients to access the HBase server and the RPC channels between the HMaster and RegionServer, are encrypted. If this parameter is set to **privacy**, the channels are encrypted and the authentication, integrity, and privacy functions are enabled. If this parameter is set to **integrity**, the channels are not encrypted and only the authentication and integrity functions are enabled. If this parameter is set to **authentication**, the channels are not encrypted, only packets are authenticated, and integrity and privacy are not required. **NOTE** The privacy mode encrypts transmitted content, including sensitive information such as user tokens, to ensure the security of the transmitted content. However, this mode has great impact on performance. Compared with the other two modes, this mode reduces read/write performance by about 60%. Modify the configuration based on the enterprise security requirements. The configuration items on the client and server must be the same.	-
HDFS	dfs.encrypt.data.transfer	Indicates whether the HDFS data transfer channels and the channels for clients to access HDFS are encrypted. The HDFS data transfer channels include the data transfer channels between DataNodes and the Data Transfer (DT) channels for clients to access DataNodes. The value **true** indicates that the channels are encrypted. The channels are not encrypted by default.	false

Service	Parameter	Description	Default Value
HDFS	dfs.encrypt.data.transfer.algorithm	Indicates whether the HDFS data transfer channels and the channels for clients to access HDFS are encrypted. This parameter is valid only when **dfs.encrypt.data.transfer** is set to **true**. The default value is **3des**, indicating that 3DES algorithm is used to encrypt data. The value can also be set to **rc4**. However, to avoid security risks, you are not advised to set the parameter to this value.	3des
HDFS	hadoop.rpc.protection	Indicates whether the RPC channels of each module in Hadoop are encrypted. The channels include: ● RPC channels for clients to access HDFS ● RPC channels between modules in HDFS, for example, between DataNode and NameNode ● RPC channels for clients to access YARN ● RPC channels between NodeManager and ResourceManager ● RPC channels for Spark to access YARN and HDFS ● RPC channels for MapReduce to access YARN and HDFS ● RPC channels for HBase to access HDFS The default value is **privacy**, indicating encrypted transmission. The value **authentication** indicates that transmission is not encrypted. **NOTE** You can set this parameter on the HDFS component configuration page. The parameter setting is valid globally, that is, the setting of whether the RPC channel is encrypted takes effect on all modules in Hadoop.	● Security mode: **privacy** ● Normal mode: **authentication**

Setting the Maximum Number of Concurrent Web Connections

To ensure web server reliability, new connections are rejected when the number of user connections reaches a specific threshold. This prevents DDOS attacks and service unavailability caused by too many users accessing the web server at the same time.

Page access for setting parameters: On FusionInsight Manager, click **Cluster**, click the name of the desired cluster, click **Services**, and click the target service. On the displayed page, click **Configuration** and click **All Configurations**. Enter a parameter name in the search box.

Table 9-85 Parameter description

Service	Parameter	Description	Default Value
HDFS/Yarn	hadoop.http.server.MaxRequests	Specifies the maximum number of concurrent web connections of each component.	2000
Spark2x	spark.connection.maxRequest	Specifies the maximum number of request connections of JobHistory.	5000

9.12.3.5 Configuring an IP Address Whitelist for Modification Allowed by HBase

If the Replication function is enabled for HBase clusters, a protection mechanism for data modification is added on the standby HBase cluster to ensure data consistency between the active and standby clusters. Upon receiving an RPC request for data modification, the standby HBase cluster checks the permission of the user who sends the request (only HBase manage users have the modification permission). Then it checks the validity of the source IP address of the request. Only modification requests from IP addresses in the white list are accepted. The IP address white list is configured by the **hbase.replication.allowedIPs** item.

Log in to FusionInsight Manager and choose **Cluster** > **Services** > **HBase**. Click **Configurations** and enter the parameter name in the search box.

Table 9-86 Parameter description

Parameter	Description	Default Value
hbase.replication.allowedIPs	Allows replication request processing from configured IP addresses only. It supports comma separated regex patterns. Each pattern can be any of the following: ● Regex pattern Example: 10.18.40.*, 10.18.*, 10.18.40.11 ● Range pattern (Range can be specified only in the last octet) Example: 10.18.40.[10-20] If this item is empty (default value), the white list contains only the IP address of the RegionServer of the cluster, indicating that only modification requests from the RegionServer of the standby HBase cluster are accepted.	N/A

9.12.3.6 Updating a Key for a Cluster

Scenario

When a cluster is installed, an encryption key is generated automatically by the system so that the security information in the cluster (such as all database user passwords and key file access passwords) can be stored in encryption mode. After the cluster is installed, if the original key is accidentally disclosed or a new key is required, you can manually update the key.

Impact on the System

- After a cluster key is updated, a new key is generated randomly in the cluster. This key is used to encrypt and decrypt the newly stored data. The old key is not deleted, and it is used to decrypt data encrypted using the old key. After security information is modified, for example, a database user password is changed, the new password is encrypted using the new key.
- When a key is updated for a cluster, the cluster must be stopped and cannot be accessed.

Prerequisites

- You have obtained the IP addresses of the active and standby management nodes. For details, see **Logging In to the Management Node**.
- You have stopped the upper-layer service applications that depend on the cluster.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > *Name of the desired cluster* and click **Stop**. In the dialog box that is displayed, enter the password of the current user

and click **OK**. Wait for a while until a message indicating that the operation is successful is displayed.

Step 3 Log in to the active management node as user **omm**.

Step 4 Run the following command to disable logout upon timeout:

TMOUT=0

NOTE

> After the operations in this section are complete, run the **TMOUT=***Timeout interval* command to restore the timeout interval in a timely manner. For example, **TMOUT=600** indicates that a user is logged out if the user does not perform any operation within 600 seconds.

Step 5 Run the following command to go to the related directory:

cd ${BIGDATA_HOME}/om-server/om/tools

Step 6 Run the following command to update the cluster key:

sh updateRootKey.sh

Enter **y** as prompted.

```
The root key update is a critical operation.
Do you want to continue?(y/n):
```

If the following information is displayed, the key is updated successfully.

```
Step 4-1: The key save path is obtained successfully.
...
Step 4-4: The root key is sent successfully.
```

Step 7 On FusionInsight Manager, click **Cluster**, click the name of the desired cluster, and click **Start**.

In the displayed dialog box, click **OK**. Wait until a message is displayed, indicating that the startup is successful.

----End

9.12.3.7 Hardening the LDAP

Configuring the LDAP Firewall Policy

In the cluster adopting the dual-plane networking, the LDAP is deployed on the service plane. To ensure the LDAP data security, you are advised to configure the firewall policy in the cluster to disable relevant LDAP ports.

Step 1 Log in to FusionInsight Manager.

Step 2 Click **Cluster**, click the name of the desired cluster, choose **Services** > **LdapServer**, and click **Configurations**.

Step 3 Check the value of **LDAP_SERVER_PORT**, which is the service port of LdapServer.

Step 4 To ensure data security, configure the firewall policy for the whole cluster to disable the LdapServer port based on the customer's firewall environment.

----End

Enabling the LDAP Audit Log Output

Users can set the audit log output level of the LDAP service and output audit logs in a specified directory, for example, **/var/log/messages**. The logs output can be used to check user activities and operation commands.

NOTE

If the function of LDAP audit log output is enabled, massive logs are generated, affecting the cluster performance. Exercise caution when enabling this function.

Step 1 Log in to any LdapServer node.

Step 2 Run the following command to edit the **slapd.conf.consumer** file, and set the value of **loglevel** to **256** (you can run the **man slapd.conf** command on the OS to view the log level definition).

cd ${BIGDATA_HOME}/FusionInsight_BASE_8.1.0.1/install/FusionInsight-ldapserver-2.7.0/ldapserver/local/template

vi slapd.conf.consumer

```
...
pidfile         [PID_FILE_SLAPD_PID]
argsfile        [PID_FILE_SLAPD_ARGS]
loglevel   256
...
```

Step 3 Log in to FusionInsight Manager, click **Cluster**, click the name of the desired cluster, choose **Services** > **LdapServer**. On the displayed page, choose **More** > **Restart Service**. Enter the administrator password and restart the service.

----End

9.12.3.8 Configuring Kafka Data Encryption During Transmission

Scenario

Data between the Kafka client and the broker is transmitted in plain text. The Kafka client may be deployed in an untrusted network, exposing the transmitting data to leakage and tampering risks.

Procedure

The channel between components is not encrypted by default. You can set the following parameters to enable security channel encryption.

Page access for setting parameters: On FusionInsight Manager, click **Cluster**, click the name of the desired cluster, and choose **Services** > **Kafka**. On the displayed page, click **Configuration** and click **All Configurations**. Enter a parameter name in the search box.

NOTE

After the configuration, restart the corresponding service for the settings to take effect.

Table 9-87 describes the parameters related to transmission encryption on the Kafka server.

Table 9-87 Parameters relevant to Kafka data encryption during transmission

Parameter	Description	Default Value
ssl.mode.enable	Indicates whether to enable the Secure Sockets Layer (SSL) protocol. If this parameter is set to **true**, services relevant to the SSL protocol are started during the broker startup.	false
security.inter.broker.protocol	Indicates communication protocol between brokers. The communication protocol can be PLAINTEXT, SSL, SASL_PLAINTEXT, or SASL_SSL.	SASL_PLAINTEXT

The SSL protocol can be configured for the server or client to encrypt transmission and communication only after **ssl.mode.enable** is set to **true** and broker enables the **SSL** and **SASL_SSL** protocols.

9.12.3.9 Configuring HDFS Data Encryption During Transmission

Configuring HDFS Security Channel Encryption

The channel between components is not encrypted by default. You can set parameters to enable security channel encryption.

Navigation path for setting parameters: On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS** > **Configurations**. On the displayed page, click the **All Configurations** tab. Enter a parameter name in the search box.

NOTE

After the configuration, restart the corresponding service for the settings to take effect.

Table 9-88 Parameters

Configuration Item	Description	Default Value
hadoop.rpc.protection	**NOTICE** - The setting takes effect only after the service is restarted. Rolling restart is not supported. - After the setting, you need to download the client configuration file again. Otherwise, HDFS cannot provide the read and write services. Indicates whether the RPC channels of each module in Hadoop are encrypted. The channels include: - RPC channels for clients to access HDFS - RPC channels between modules in HDFS, for example, between DataNode and NameNode - RPC channels for clients to access Yarn - RPC channels between NodeManager and ResourceManager - RPC channels for Spark to access Yarn and HDFS - RPC channels for MapReduce to access Yarn and HDFS - RPC channels for HBase to access HDFS **NOTE** The setting takes effect globally, that is, the encryption attribute of the RPC channel of each module in the Hadoop takes effect.	- Security mode: privacy - Normal mode: authentication **NOTE** - **authentication**: indicates that only authentication is required. - **integrity**: indicates that authentication and consistency check need to be performed. - **privacy**: indicates that authentication, consistency check, and encryption need to be performed.

Configuration Item	Description	Default Value
dfs.encrypt.data.transfer	Indicates whether the HDFS data transfer channels and the channels for clients to access HDFS are encrypted. The HDFS data transfer channels include the data transfer channels between DataNodes and the Data Transfer (DT) channels for clients to access DataNodes. The value **true** indicates that the channels are encrypted. The channels are not encrypted by default. **NOTE** ● This parameter is valid only when **hadoop.rpc.protection** is set to **privacy**. ● If a large amount of service data is transmitted, enabling encryption by default severely affects system performance. ● If data transmission encryption is configured for one cluster in the trusted cluster, the same data transmission encryption must be configured for the peer cluster.	false
dfs.encrypt.data.transfer.algorithm	Indicates the algorithm to encrypt the HDFS data transfer channels and the channels for clients to access HDFS. This parameter is valid only when **dfs.encrypt.data.transfer** is set to **true**. **NOTE** The default value is **3des**, indicating that 3DES algorithm is used to encrypt data. The value can also be set to **rc4**. However, to avoid security risks, you are not advised to set the parameter to this value.	3des
dfs.encrypt.data.transfer.cipher.suites	This parameter can be left empty or set to **AES/CTR/NoPadding** to specify the cipher suite for data encryption. If this parameter is not specified, the encryption algorithm specified by **dfs.encrypt.data.transfer.algorithm** is used for data encryption. The default value is **AES/CTR/NoPadding**.	AES/CTR/NoPadding

9.12.3.10 Encrypting the Communication Between the Controller and the Agent

Scenario

After a cluster is installed, Controller and Agent need to communicate with each other. The Kerberos authentication is used during the communication. By default, the communication is not encrypted during the communication for the sake of cluster performance. Users who have demanding security requirements can use the method described in this section for encryption.

Impact on the System

- Controller and all Agents automatically restart, which interrupts FusionInsight Manager.
- The performance of management nodes deteriorates in large clusters. You are advised to enable the encryption function for clusters with a maximum of 200 nodes.

Prerequisites

You have obtained the IP addresses of the active and standby management nodes.

Procedure

Step 1 Log in to the active management node as user **omm**.

Step 2 Run the following command to disable logout upon timeout:

TMOUT=0

NOTE

After the operations in this section are complete, run the **TMOUT=***Timeout interval* command to restore the timeout interval in a timely manner. For example, **TMOUT=600** indicates that a user is logged out if the user does not perform any operation within 600 seconds.

Step 3 Run the following command to go to the related directory:

cd ${CONTROLLER_HOME}/sbin

Step 4 Run the following command to enable communication encryption:

./enableRPCEncrypt.sh -t

Run the **sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh** command to check whether **ResHAStatus** of the active management node Controller is **Normal** and whether you can log in to FusionInsight Manager again. If yes, the enablement is successful.

Step 5 Run the following command to disable communication encryption when necessary:

./enableRPCEncrypt.sh -f

Run the **sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh** command to check whether **ResHAStatus** of the active management node Controller is **Normal** and whether you can log in to FusionInsight Manager again. If yes, the enablement is successful.

----End

9.12.3.11 Updating SSH Keys for User omm

Scenario

During cluster installation, the system automatically generate the SSH public key and private key for user **omm** to establish the trust relationship between nodes. After the cluster is installed, if the original keys are accidentally disclosed or new keys are used, the system administrator can perform the following operations to manually change the keys.

Prerequisites

- The cluster has been stopped.
- No other management operations are being performed.

Procedure

Step 1 Log in as user **omm** to the node whose SSH keys need to be replaced.

If the node is a Manager management node, run the following command on the active management node.

Step 2 Run the following command to disable logout upon timeout:

TMOUT=0

NOTE

After the operations in this section are complete, run the **TMOUT=***Timeout interval* command to restore the timeout interval in a timely manner. For example, **TMOUT=600** indicates that a user is logged out if the user does not perform any operation within 600 seconds.

Step 3 Run the following command to generate a key for the node:

- If the node is a Manager management node, run the following command:
  sh ${CONTROLLER_HOME}/sbin/update-ssh-key.sh
- If the node is a non-Manager management node, run the following command:
  sh ${NODE_AGENT_HOME}/bin/update-ssh-key.sh

If "Succeed to update ssh private key." is displayed when the preceding command is executed, the SSH key is generated successfully.

Step 4 Run the following command to copy the public key of the node to the active management node:

scp ${HOME}/.ssh/id_rsa.pub** *oms_ip*:**${HOME}/.ssh/id_rsa.pub_bak

oms_ip: indicates the IP address of the active management node.

Enter the password of user **omm** to copy the files.

Step 5 Log in to the active management node as user **omm**.

Step 6 Run the following command to disable logout on system timeout:

```
TMOUT=0
```

Step 7 Run the following command to go to the related directory:

```
cd ${HOME}/.ssh
```

Step 8 Run the following command to add new public keys:

```
cat id_rsa.pub_bak >> authorized_keys
```

Step 9 Run the following command to move the temporary public key file, for example, **/tmp**.

```
mv -f id_rsa.pub_bak /tmp
```

Step 10 Copy the **authorized_keys** file of the active management node to the other nodes in the cluster:

```
scp authorized_keys node_ip:/${HOME}/.ssh/authorized_keys
```

node_ip: indicates the IP address of another node in the cluster. Multiple IP addresses are not supported.

Step 11 Run the following command to confirm private key replacement without entering the password:

```
ssh node_ip
```

node_ip: indicates the IP address of another node in the cluster. Multiple IP addresses are not supported.

Step 12 Log in to FusionInsight Manager. On **Homepage**, locate the desired cluster and choose ... > **Start** to start the cluster.

----End

9.12.4 Security Maintenance

9.12.4.1 Account Maintenance Suggestions

It is recommended that the administrator conduct routine checks on the accounts. The check covers the following items:

- Check whether the accounts of the OS, FusionInsight Manager, and each component are necessary and whether temporary accounts have been deleted.
- Check whether the permissions of the accounts are appropriate. Different administrators have different rights.
- Check and audit the logins and operation records of all types of accounts.

9.12.4.2 Password Maintenance Suggestions

User identity authentication is a must for accessing the application system. The complexity and validity period of user accounts and passwords must meet customers' security requirements.

The password maintenance suggestions are as follows:

1. Dedicated personnel must be arranged to manage the OS password.
2. The passwords must meet the complexity requirements, such as minimum password length or character types.
3. Passwords must be encrypted before transfer. Generally, do not transfer passwords using emails.
4. Passwords must be encrypted in configuration files.
5. Enterprise users need to change the passwords when the system is handed over.
6. Passwords must be periodically changed.

9.12.4.3 Log Maintenance Suggestions

Operation logs help discover exceptions such as illegal operations and login by unauthorized users. The system records important operations in logs. You can use operation logs to locate problems.

Checking Logs Regularly

Check system logs periodically and handle exceptions such as unauthorized operations or logins in a timely manner.

Backing Up Logs Regularly

The audit logs provided by FusionInsight Manager and cluster record the user activities and operations. You can export the audit logs on FusionInsight Manager. If there are too many audit logs in the system, you can configure dump parameters to dump audit logs to a specified server to ensure that the cluster nodes disk space is sufficient.

Maintenance Owner

Network monitoring engineers and system maintenance engineers

9.12.5 Security Statement

JDK Usage Statement

MRS MRS cluster is a big data cluster that provides users with distributed data analysis and computing capabilities. The built-in JDK of MRS MRS is OpenJDK, which is used in the following scenarios:

- Platform service running and maintenance
- Linux client operations, including service submission and application O&M

JDK Risk Description

The system performs permission control on the built-in JDK. Only users in the related group of the FusionInsight platform can access the JDK. In addition, the platform is deployed on a customer's intranet. Therefore, the security risk is low.

JDK Hardening

For details about how to harden the JDK, see "Hardening JDK" in **Hardening Policies**.

Public IP Addresses in Hue

Hue uses the test cases of third-party packages, such as **ipadrress**, **requests**, and **Django**, and uses the public IP addresses in the comments of the test cases. However, these public IP addresses are not involved when Hue provides services, and the Hue configuration file does not involve these public IP addresses.

9.13 Alarm Reference (Applicable to MRS 3.*x*)

9.13.1 ALM-12001 Audit Log Dumping Failure

Description

Cluster audit logs need to be dumped on a third-party server due to the local historical data backup policy. The system starts to check the dump server at 3 a.m. every day. If the dump server meets the configuration conditions, audit logs can be successfully dumped. This alarm is generated when the audit log dump fails if the disk space of the dump directory on the third-party server is insufficient or a user changes the username, password, or dump directory of the dump server.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12001	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.

Name	Meaning
HostName	Specifies the host for which the alarm is generated.

Impact on the System

System can store a maximum of only 50 dump files locally. If the fault persists on the dump server, the local audit logs may be lost.

Possible Causes

- The network connection is abnormal.
- The username, password, or dump directory of the dump server does not meet the configuration conditions.
- The disk space of the dump directory is insufficient.

Procedure

Check whether the network connection is normal.

Step 1 On the FusionInsight Manager home page, choose **Audit > Configurations**.

Step 2 Check whether the SFTP IP on the dump configuration page is valid.

Log in to the node where Manager is located as user **root** and run the **ping** command to check whether the network connection between the SFTP server and the cluster is normal.

- If yes, go to **Step 5**.
- If no, go to **Step 3**.

Step 3 Repair the network connection, reset the SFTP password, and click **OK**.

Step 4 Wait for 2 minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check whether the username, password, or dump directory are correct.

Step 5 On the dump configuration page, check whether the username, password, and dump directory of the third-party server are correct.

- If yes, go to **Step 8**.
- If no, go to **Step 6**.

Step 6 Change the username, password, or dump directory, reset the SFTP password and click **OK**.

Step 7 Wait for 2 minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Check whether the disk space of the dump directory is sufficient.

Step 8 Log in to the third-party server as user **root** and run the **df** command to check whether the disk space of the dump directory of the third-party server exceeds 100 MB.

- If yes, go to **Step 11**.
- If no, go to **Step 9**.

Step 9 Expand disk space capacity for the third-party server, Reset the SFTP password and click **OK**

Step 10 Wait for 2 minutes, view real-time alarms and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Reset the dump rule.

Step 11 On the FusionInsight Manager home page, choose **Audit > Configurations**.

Step 12 Reset dump rules, set the parameters properly, and click **OK**.

Step 13 Wait for 2 minutes, view real-time alarms and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 14**.

Collect fault information.

Step 14 On the FusionInsight Manager, choose **O&M** > **Log > Download**.

Step 15 Select **OmmServer** from the **Service** and click **OK**.

Step 16 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 17 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.2 ALM-12004 OLdap Resource Abnormal

Description

The system checks LDAP resources every 60 seconds. This alarm is generated when the system detects that the LDAP resources in Manager are abnormal for six consecutive times.

This alarm is cleared when the Ldap resource in the Manager recovers and the alarm handling is complete.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12004	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The Manager and component WebUI authentication services are unavailable and cannot provide security authentication and user management functions for web upper-layer services. Users may be unable to log in to the WebUIs of Manager and components.

Possible Causes

The LdapServer process in the Manager is abnormal.

Procedure

Check whether the LdapServer process in the Manager is normal.

Step 1 Log in the Manager node in the cluster as user **omm**.

Log in to FusionInsight Manager using the floating IP address, and run the **sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh** command to check the information about the current Manager two-node cluster.

Step 2 Run **ps -ef | grep slapd** command to check whether the LdapServer resource process in the **${BIGDATA_HOME}/om-server/om/** in the process configuration file is running properly.

NOTE

You can determine that the resource is normal by checking the following information:

1. After the **sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh** command runs, **ResHAStatus** of the OLdap is **Normal**.
2. After the **ps -ef | grep slapd** command runs, the slapd process of port 21750 can be viewed.
   - If yes, go to **Step 3**.
   - If no, go to **Step 4**.

Step 3 Run the **kill -2** *ldap pid* command to restart the LdapServer process and wait for 20 seconds. The HA starts the OLdap process automatically. Check whether the current OLdap resource is in normal state.

- If yes, the operation is complete.
- If no, go to **Step 4**.

Collect fault information.

Step 4 On the FusionInsight Manager home page, choose **O&M** > **Log > Download**.

Step 5 Select **OmsLdapServer** and **OmmServer** from the **Service** and click **OK**.

Step 6 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 7 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.3 ALM-12005 OKerberos Resource Abnormal

Description

The alarm module checks the status of the Kerberos resource in Manager every 80 seconds. This alarm is generated when the alarm module detects that the Kerberos resources are abnormal for six consecutive times.

This alarm is cleared when the Kerberos resource recovers and the alarm handling is complete.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12005	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The component WebUI authentication services are unavailable and cannot provide security authentication functions for web upper-layer services. Users may be unable to log in to FusionInsight Manager and the WebUIs of components.

Possible Causes

The OLdap resource on which the Okerberos depends is abnormal.

Procedure

Check whether the OLdap resource on which the Okerberos depends is abnormal in the Manager.

Step 1 Log in the Manager node in the cluster as user **omm**.

Log in to FusionInsight Manager using the floating IP address, and run the **sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh** command to check the information about the current Manager two-node cluster.

Step 2 Run the **sh ${BIGDATA_HOME}/om-server/OMS/workspace0/ha/module/hacom/script/status_ha.sh** command to check whether the OLdap resource status managed by HA is normal. (In single-node mode, the OLdap resource is in the Active_normal state; in the two-node mode, the OLdap resource is in the Active_normal state on the active node and in the Standby_normal state on the standby node.)

- If yes, go to **Step 4**.
- If no, go to **Step 3**.

Step 3 See the procedure in **ALM-12004 OLdap Resource Abnormal** to resolve the problem. After the OLdap resource status recovers, check whether the OKerberos resource status is normal.

- If yes, the operation is complete.

- If no, go to **Step 4**.

Collect fault information.

Step 4 On the FusionInsight Manager home page, choose **O&M** > **Log > Download**.

Step 5 Select **OmsKerberos** and **OmmServer** from the **Service** and click **OK**.

Step 6 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 7 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.4 ALM-12006 Node Fault

Description

Controller checks the NodeAgent heartbeat every 30 seconds. If Controller does not receive heartbeat messages from a NodeAgent, it attempts to restart the NodeAgent process. This alarm is generated if the NodeAgent fails to be restarted for three consecutive times.

This alarm is cleared when Controller can properly receive the status report of the NodeAgent.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12006	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.

Name	Meaning
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

Services on the node are unavailable.

Possible Causes

The network is disconnected, the hardware is faulty, or the operating system runs slowly.

Procedure

Check whether the network is disconnected, whether the hardware is faulty, or whether the operating system runs commands slowly.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. On the page that is displayed, click ⌄ in the row containing the alarm, click the host name, and view the IP address of the host for which the alarm is generated.

Step 2 Log in to the active management node as user **root**.

Step 3 Run the **ping** *IP address of the faulty host* command to check whether the faulty node is reachable.

- If yes, go to **Step 12**.
- If no, go to **Step 4**.

Step 4 Contact the network administrator to check whether the network is faulty.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 Rectify the network fault and check whether the alarm is cleared from the alarm list.

- If yes, no further action is required.
- If no, go to **Step 6**.

Step 6 Contact the hardware administrator to check whether the hardware (CPU or memory) of the node is faulty.

- If yes, go to **Step 7**.
- If no, go to **Step 12**.

Step 7 Repair or replace faulty components and restart the node. Check whether the alarm is cleared.

- If yes, no further action is required.

- If no, go to **Step 8**.

Step 8 If a large number of node faults are reported in the cluster, the floating IP addresses may be abnormal. As a result, Controller cannot detect the NodeAgent heartbeat.

Log in to any management node and view the **/var/log/Bigdata/omm/oms/ha/scriptlog/floatip.log** log to check whether the logs generated one to two minutes before and after the faults occur are complete.

For example, a complete log is in the following format:

```
2017-12-09 04:10:51,000 INFO (floatip) Read from ${BIGDATA_HOME}/om-server_*/om/etc/om/
routeSetConf.ini,value is : yes
2017-12-09 04:10:51,000 INFO (floatip) check wsNetExport : eth0 is up.
2017-12-09 04:10:51,000 INFO (floatip) check omNetExport : eth0 is up.
2017-12-09 04:10:51,000 INFO (floatip) check wsInterface : eRth0:oms, wsFloatIp: XXX.XXX.XXX.XXX.
2017-12-09 04:10:51,000 INFO (floatip) check omInterface : eth0:oms, omFloatIp: XXX.XXX.XXX.XXX.
2017-12-09 04:10:51,000 INFO (floatip) check  wsFloatIp : XXX.XXX.XXX.XXX is reachable.
2017-12-09 04:10:52,000 INFO (floatip) check  omFloatIp : XXX.XXX.XXX.XXX is reachable.
```

- If yes, go to **Step 12**.
- If no, go to **Step 9**.

Step 9 Check whether the omNetExport log is printed after the wsNetExport is detected or whether the interval for printing two logs exceeds 10 seconds or longer.

- If yes, go to **Step 10**.
- If no, go to **Step 12**.

Step 10 View the **/var/log/message** file of the OS to check whether sssd frequently restarts or nscd exception information is displayed when the fault occurs. For Red Hat, check sssd information. For SUSE, check nscd information.

sssd restart example

```
Feb  7 11:38:16 10-132-190-105 sssd[pam]: Shutting down
Feb  7 11:38:16 10-132-190-105 sssd[nss]: Shutting down
Feb  7 11:38:16 10-132-190-105 sssd[nss]: Shutting down
Feb  7 11:38:16 10-132-190-105 sssd[be[default]]: Shutting down
Feb  7 11:38:16 10-132-190-105 sssd: Starting up
Feb  7 11:38:16 10-132-190-105 sssd[be[default]]: Starting up
Feb  7 11:38:16 10-132-190-105 sssd[nss]: Starting up
Feb  7 11:38:16 10-132-190-105 sssd[pam]: Starting up
```

Example nscd exception information

```
Feb 11 11:44:42 10-120-205-33 nscd: nss_ldap: failed to bind to LDAP server ldaps://10.120.205.55:21780:
Can't contact LDAP server
Feb 11 11:44:43 10-120-205-33 ntpq: nss_ldap: failed to bind to LDAP server ldaps://10.120.205.55:21780:
Can't contact LDAP server
Feb 11 11:44:44 10-120-205-33 ntpq: nss_ldap: failed to bind to LDAP server ldaps://10.120.205.92:21780:
Can't contact LDAP server
```

- If yes, go to **Step 11**.
- If no, go to **Step 12**.

Step 11 Check whether the LdapServer node is faulty, for example, the service IP address is unreachable or the network latency is too high. If the fault occurs periodically, locate and eliminate it and run the **top** command to check whether abnormal software exists.

Collect the fault information.

Step 12 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 13 Select the following nodes from **Services** and click **OK**.

- NodeAgent
- Controller
- OS

Step 14 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 15 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.5 ALM-12007 Process Fault

Description

This alarm is generated when the process health check module detects that the process connection status is **Bad** for three consecutive times. The process health check module checks the process status every 5 seconds.

This alarm is cleared when the process can be connected.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12007	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.

Name	Meaning
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The service provided by the process is unavailable.

Possible Causes

- The instance process is abnormal.
- The disk space is insufficient.

NOTE

If a large number of process fault alarms exist in a time segment, files in the installation directory may be deleted mistakenly or permission on the directory may be modified.

Procedure

Check whether the instance process is abnormal.

Step 1 In the FusionInsight Manager portal, click **O&M** > **Alarm** > **Alarms**, click ⌄ in the row where the alarm is located , and click the host name to view the host address for which the alarm is generated

Step 2 On the **Alarms** page, check whether the **ALM-12006 Node Fault** is generated.

- If yes, go to **Step 3**.
- If no, go to **Step 4**.

Step 3 Handle the alarm according to **ALM-12006 Node Fault**.

Step 4 Log in to the host for which the alarm is generated as user **root**. Check whether the installation directory user, user group, and permission of the alarm role are correct. The user, user group, and the permission must be **omm:ficommon 750**.

For example, the NameNode installation directory is *${BIGDATA_HOME}*/**FusionInsight_Current/***1_8_NameNode***/etc**.

- If yes, go to **Step 6**.
- If no, go to **Step 5**.

Step 5 Run the following command to set the permission to **750** and **User:Group** to **omm:ficommon**:

chmod 750 *<folder_name>*

chown omm:ficommon *<folder_name>*

Step 6 Wait for 5 minutes. In the alarm list, check whether **ALM-12007 Process Fault** is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Check whether disk space is sufficient.

Step 7 On the FusionInsight Manager, check whether the alarm list contains **ALM-12017 Insufficient Disk Capacity**.

- If yes, go to **Step 8**.
- If no, go to **Step 11**.

Step 8 Rectify the fault by following the steps provided in **ALM-12017 Insufficient Disk Capacity**.

Step 9 Wait for 5 minutes. In the alarm list, check whether **ALM-12017 Insufficient Disk Capacity** is cleared.

- If yes, go to **Step 10**.
- If no, go to **Step 11**.

Step 10 Wait for 5 minutes. In the alarm list, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Collect fault information.

Step 11 On the FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 12 According to the service name obtained in **Step 1**, select the component and **NodeAgent** from the **Service** and click **OK**.

Step 13 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 14 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.6 ALM-12010 Manager Heartbeat Interruption Between the Active and Standby Nodes

Description

This alarm is generated when the active Mager does not receive the heartbeat signal from the standby Manager within 7 seconds.

This alarm is cleared when the active Manager receives heartbeat signals from the standby Manager.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12010	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

When the active Manager process is abnormal, an active/standby failover cannot be performed, and services are affected.

Possible Causes

- The link between the active and standby Manager is abnormal.
- The node name configuration is incorrect.
- The port is disabled by the firewall.

Procedure

Check whether the network between the active and standby Manager server is normal.

Step 1 In the FusionInsight Manager portal, click **O&M** > **Alarm** > **Alarms**, click ⌄ in the row containing the alarm and view the IP address of the standby Manager (Peer Manager) server in the alarm details.

Step 2 Log in to the active Manager server as user **root**.

Step 3 Run the **ping** *standby Manager heartbeat IP address* command to check whether the standby Manager server is reachable.

- If yes, go to **Step 6**.
- If no, go to **Step 4**.

Step 4 Contact the network administrator to check whether the network is faulty.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 Rectify the network fault and check whether the alarm is cleared from the alarm list.

- If yes, no further action is required.
- If no, go to **Step 6**.

Step 6 Run the following command to go to the software installation directory:

cd /opt

Step 7 Run the following command to find the configuration file directory of the active and standby nodes.

find -name hacom_local.xml

Step 8 Run the following command to go to the **workspace** directory:

cd${BIGDATA_HOME}/om-server/OMS/workspace0/ha/local/hacom/conf/

Step 9 Run the **vim** command to open the **hacom_local.xml** file. Check whether the local and peer nodes are correctly configured. The local node is configured as the active node, and the peer node is configured as the standby node.

- If yes, go to **Step 12**.
- If no, go to **Step 10**.

Step 10 Modify the configuration of the active and standby nodes in the **hacom_local.xml** file and press **Esc** to return to the command mode. Run the **:wq** command to save the modification and exit.

Step 11 Check whether the alarm is cleared automatically.

- If yes, no further action is required.
- If no, go to **Step 12**.

Check whether the port is disabled by the firewall.

Step 12 Run the **lsof -i :20012** command to check whether the heartbeat ports of the active and standby nodes are enabled. If the command output is displayed, the ports are enabled. Otherwise, the ports are disabled by the firewall.

- If yes, go to **Step 13**.
- If no, go to **Step 16**.

Step 13 Run the **iptables -P INPUT ACCEPT** command to avoid the server disconnection.

Step 14 Run the following command to clear the firewall:

iptables -F

Step 15 Check whether the alarm is cleared from the alarm list.

- If yes, no further action is required.

- If no, go to **Step 16**.

Collect fault information.

Step 16 On the FusionInsight Manager, choose **O&M** > **Log > Download**.

Step 17 Select the following nodes from the **Service** and click **OK**:

- OmmServer
- Controller
- NodeAgent

Step 18 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 19 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.7 ALM-12011 Manager Data Synchronization Exception Between the Active and Standby Nodes

Description

The system checks data synchronization between the active and standby Manager nodes every 60 seconds. This alarm is generated when the standby Manager fails to synchronize files with the active Manager.

This alarm is cleared when the standby Manager synchronizes files with the active Manager.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12011	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

Some configurations will be lost after an active/standby switchover because the configuration files on the standby Manager are not updated. Maybe Manager and some components cannot run properly.

Possible Causes

- The link between the active and standby Managers is interrupted or The storage space of the **/srv/BigData/LocalBackup** directory is full.
- The synchronization file does not exist or the file permission is incorrect.

Procedure

Check whether the network between the active Manager server and the standby Manager server is normal.

Step 1 In the FusionInsight Manager portal, click **O&M** > **Alarm** > **Alarms**, click ⌄ in the row where the alarm is located and obtain the standby Manager server IP address (Peer Manager IP address) in the alarm details.

Step 2 Log in to the active Manager server as user **root**.

Step 3 Run the **ping** *standby Manager IP address* command to check whether the standby Manager server is reachable.

- If yes, go to **Step 6**.
- If no, go to **Step 4**.

Step 4 Contact the network administrator to check whether the network is faulty.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 Rectify the network fault and check whether the alarm is cleared from the alarm list.

- If yes, no further action is required.
- If no, go to **Step 6**.

Check whether the storage space of the /srv/BigData/LocalBackup directory is full.

Step 6 Run the following command to check whether the storage space of the **/srv/BigData/LocalBackup** directory is full:

df -hl /srv/BigData/LocalBackup

- If yes, go to **Step 7**.
- If no, go to **Step 10**.

Step 7 Run the following command to clear unnecessary backup files:

rm -rf *Directory to be cleared*

Example:

rm -rf /srv/BigData/LocalBackup/0/default-oms_20191211143443

Step 8 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

In the **Operation** column of the backup task to be performed, click **Configure** and change the value of **Maximum Number of Backup Copies** to reduce the number of backup file sets.

Step 9 Wait about 1 minute and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 10**.

Check whether the synchronization file exists and whether the file permission is normal.

Step 10 Run the following command to check whether the synchronization file exists.

find /srv/BigData/ -name "sed*"

find /opt -name "sed*"

- If yes, go to **Step 11**.
- If no, go to **Step 12**.

Step 11 Run the following command to view the synchronization file information and permission obtained in **Step 10**.

ll *path of the file to be found*

- If the size of the file is 0 and the permission column is **-**, the file is a junk file. Run the following command to delete it.

  rm -rf *files to be deleted*

  Wait for several minutes and check whether the alarm is cleared. If the alarm persists, go to **Step 12**.
- If the file size is not 0, go to **Step 12**.

Step 12 View the log files generated when the alarm is generated.

1. Run the following command to switch to the HA run log file path.

   cd /var/log/Bigdata/omm/oms/ha/runlog/
2. Decompress and view the log files generated when the alarm is generated.

   For example, if the name of the file to be viewed is **ha.log.2021-03-22_12-00-07.gz**, run the following command:

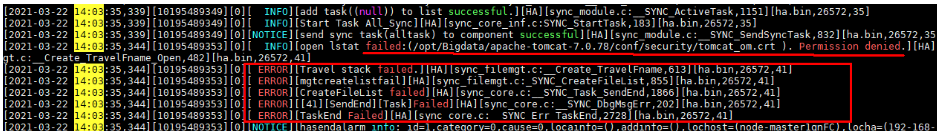
   gunzip *ha.log.2021-03-22_12-00-07.gz*

vi *ha.log.2021-03-22_12-00-07*

Check whether error information is reported before and after the alarm generation time.

- If yes, rectify the fault based on the error information. Then go to **Step 13**.

  For example, if the following error information is displayed, the directory permission is insufficient. In this case, change the directory permission to be the same as that on the normal node.



```
[2021-03-22 14:03:35,339][10195489349][0][  INFO][add task((null)) to list successful.][HA][sync_module.c:__SYNC_ActiveTask,1151][ha.bin,26572,35]
[2021-03-22 14:03:35,339][10195489349][0][  INFO][Start Task All_Sync][HA][sync_core_inf.c:SYNC_StartTask,183][ha.bin,26572,35]
[2021-03-22 14:03:35,339][10195489349][0][NOTICE][send sync task(alltask) to component successful][HA][sync_module.c:__SYNC_SendSyncTask,832][ha.bin,26572,35
[2021-03-22 14:03:35,344][10195489353][0][  INFO][open lstat failed:(/opt/Bigdata/apache-tomcat-7.0.78/conf/security/tomcat_om.crt ). Permission denied.][HA]
gt.c:__Create_TravelFname_Open,482][ha.bin,26572,41]
[2021-03-22 14:03:35,344][10195489353][0][ ERROR][Travel stack failed.][HA][sync_filemgt.c:__Create_TravelFname,613][ha.bin,26572,41]
[2021-03-22 14:03:35,344][10195489353][0][ ERROR][mgtcreatelistfail][HA][sync_filemgt.c:_SYNC_CreateFileList,855][ha.bin,26572,41]
[2021-03-22 14:03:35,344][10195489353][0][ ERROR][CreateFileList failed][HA][sync_core.c:__SYNC_Task_SendEnd,1866][ha.bin,26572,41]
[2021-03-22 14:03:35,344][10195489353][0][ ERROR][[41][SendEnd][Task]Failed][HA][sync_core.c:__SYNC_DbgMsgErr,202][ha.bin,26572,41]
[2021-03-22 14:03:35,344][10195489353][0][ ERROR][TaskEnd Failed][HA][sync core.c:  SYNC Err TaskEnd,2728][ha.bin,26572,41]
[2021-03-22 14:03:35,344][10195489353][0][NOTICE][hasendalarm info: id=1,category=0,cause=0,locainfo=(),addinfo=(),lochost=(node-master1gnFC),locha=(192-168-
```

- If no, go to **Step 14**.

Step 13 Wait about 10 minute and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 14**.

Collect fault information.

Step 14 On the FusionInsight Manager, choose **O&M** > **Log > Download**.

Step 15 Select the following nodes from the **Service** and click **OK**:

- OmmServer
- Controller
- NodeAgent

Step 16 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 17 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.8 ALM-12012 NTP Service Is Abnormal

Description

The system checks whether the NTP service on a node synchronizes time with the NTP service on the active OMS node every 60 seconds. This alarm is generated when the NTP service fails to synchronize time for two consecutive times.

This alarm is generated when the time difference between the NTP service on a node and the NTP service on the active OMS node is greater than or equal to 20s for two consecutive times. This alarm is cleared when the time difference is less than 20s.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12012	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The time on the node is inconsistent with that on other nodes in the cluster. Therefore, some FusionInsight applications on the node may not run properly.

Possible Causes

- The NTP service on the current node cannot start properly.
- The current node fails to synchronize time with the NTP service on the active OMS node.
- The key value authenticated by the NTP service on the current node is inconsistent with that on the active OMS node.
- The time offset between the node and the NTP service on the active OMS node is large.

Procedure

Check the NTP service mode of the node.

Step 1 Log in to the active management node as user **root**, run the **su - omm** command to switch to user **omm**, and run the following command to check the resource status on the active and standby nodes:

sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh

- If "chrony" is displayed in the **ResName** column of the command output, go to **Step 2**.
- If "ntp" is displayed in the **ResName** column, go to **Step 20**.

NOTE

If both "chrony" and "ntp" are displayed in the **ResName** column of the command output, the NTP service mode is being switched. Wait for 10 minutes and go to **Step 1** again. If both "chrony" and "ntp" persist, contact O&M personnel personnel.

Check whether the chrony service on the node is started properly.

Step 2 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. On the page that is displayed, click ∨ in the row containing the alarm, and view the name of the host for which the alarm is generated in **Location**.

Step 3 Check whether the chronyd process is running on the node where the alarm is generated. Log in to the node for which the alarm is generated as user **root** and run the **ps -ef | grep chronyd | grep -v grep** command to check whether the command output contains the chronyd process.

- If yes, go to **Step 6**.
- If no, go to **Step 4**.

Step 4 Run the **systemctl chronyd start** command to start the NTP service. (Currently, only CentOS and Red Hat Enterprise Linux 7.0 or later are supported.)

Step 5 Check whether the alarm is cleared 10 minutes later.

- If yes, no further action is required.
- If no, go to **Step 6**.

Check whether the current node can synchronize time properly with the chrony service on the active OMS node.

Step 6 Check whether the node can synchronize time with the NTP service on the active OMS node based on additional information of the alarm.

- If yes, go to **Step 7**.
- If no, go to **Step 17**.

Step 7 Check whether the synchronization with the chrony service on the active OMS node is faulty.

Log in to the node for which the alarm is generated as user **root** and run the **chronyc sources** command.

In the command output, if there is an asterisk (*) before the IP address of the chrony service on the active OMS node, the synchronization is normal. The command output is as follows:

```
MS Name/IP address         Stratum Poll Reach LastRx Last sample
===============================================================================
^* 10.10.10.162            10  10   377   626    +16us[  +15us] +/-  308us
```

In the command output, if there is no asterisk (*) before the IP address of the NTP service on the active OMS node, and the value of **Reach** is **0**, the synchronization is abnormal.

```
MS Name/IP address         Stratum Poll Reach LastRx Last sample
===============================================================================
^? 10.1.1.1                      0  10     0     -     +0ns[   +0ns] +/-    0ns
```

- If yes, go to **Step 8**.
- If no, go to **Step 38**.

Step 8 The chrony synchronization failure is typically caused by the system firewall. If the firewall can be disabled, disable it. If the firewall cannot be disabled, check the firewall configuration policy and ensure that UDP ports 123 and 323 are not disabled. (For details, see the firewall configuration policy of each system.)

Step 9 Check whether the alarm is cleared 10 minutes later.

- If yes, no further action is required.
- If no, go to **Step 10**.

Step 10 Log in to the active OMS node as user **root** and run the following command to view the authentication code whose key value index is **1M**:

In Red Hat Enterprise Linux, run the **cat ${BIGDATA_HOME}/om-server/OMS/workspace/conf/chrony.keys** command.

Step 11 Run the following command to check whether the key value is the same as that queried in **Step 10**:

In Red Hat Enterprise Linux, run the **diff ${BIGDATA_HOME}/om-server/OMS/workspace/conf/chrony.keys /etc/chrony.keys** command.

NOTE

If the key values are the same, no result is returned after the command is executed. For example:

```
host01:~ # cat ${BIGDATA_HOME}/om-server/OMS/workspace/conf/chrony.keys
1 M sdYbq;o^CzEAWo<U=Tw5
host01:~ # diff ${BIGDATA_HOME}/om-server/OMS/workspace/conf/chrony.keys /etc/chrony.keys
host01:~ #
```

- If yes, go to **Step 12**.
- If no, go to **Step 38**.

Step 12 Run the **cat ${BIGDATA_HOME}/om-server/om/packaged-distributables/ntpKeyFile** command to check whether the key value is the same as that queried in **Step 10**. (Compare the key value with that of the authentication key index field **1M** queried in **Step 10**.)

- If yes, go to **Step 13**.
- If no, go to **Step 15**.

Step 13 Log in to the faulty node as user **root** and run the **cat /etc/chrony.keys** command in Red Hat Enterprise Linux to check whether the key value is the same as the value queried in **Step 12** (use the key value of the authentication key index field **1M** for comparison).

- If yes, go to **Step 38**.
- If no, go to **Step 14**.

Step 14 Run the **su - omm** command to switch to user **omm**, change the key value of the authentication key index field **1M** in **${NODE_AGENT_HOME}/chrony.keys** to the key value of **ntpKeyFile** in **Step 12**, and go to **Step 16**.

Step 15 Run the following commands as user **root** or **omm** to change the NTP key value of the active OMS node (change **ntp.keys** to **ntpkeys** in Red Hat Enterprise Linux):

```
cd ${BIGDATA_HOME}/om-server/OMS/workspace/conf
```

```
sed -i "`cat chrony.keys | grep -n '1 M'|awk -F ':' '{print $1}'`d" chrony.keys
```

```
echo "1 M `cat ${BIGDATA_HOME}/om-server/om/packaged-distributables/ntpKeyFile`" >> chrony.keys
```

Check whether the key value of the authentication key index field **1M** in **chrony.keys** is the same as that of **ntpKeyFile**.

- If yes, go to **Step 16**.
- If no, change the key value of the authentication key index field **1M** in **chrony.keys** to the key value of **ntpKeyFile** and go to **Step 16**.

Step 16 After 5 minutes, run the **systemctl chronyd restart** command to restart the chrony service on the active OMS node. After 15 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 38**.

Check whether the time deviation between the node and the chrony service on the active OMS node is large.

Step 17 Check whether the time deviation is large in additional information of the alarm.

- If yes, go to **Step 18**.
- If no, go to **Step 38**.

Step 18 On the **Hosts** tab page, select the host for which the alarm is generated, and choose **More** > **Stop All Instances** to stop all the services on the node.

If the time on the alarm node is later than that on the chrony service of the active OMS node, adjust the time of the alarm node. After adjusting the time, choose **More** > **Start All Instances** to start the services on the node.

If the time on the alarm node is earlier than that on the chrony service of the active OMS node, wait until the time deviation is due and adjust the time of the alarm node. After adjusting the time, choose **More** > **Start All Instances** to start the services on the node.

NOTE

If you do not wait, data loss may occur.

Step 19 After 10 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 38**.

Check whether the NTP service on the node is started properly.

Step 20 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. On the page that is displayed, click ⌄ in the row containing the alarm, and view the name of the host for which the alarm is generated in **Location**.

Step 21 Check whether the ntpd process is running on the node using the following method. Log in to the alarm node as user **root** and run the **ps -ef | grep ntpd | grep -v grep** command to check whether the command output contains the ntpd process.

- If yes, go to **Step 24**.
- If no, go to **Step 22**.

Step 22 Run the **service ntp start** command (or the **service ntpd start** command in Red Hat Enterprise Linux) to start the NTP service.

Step 23 After 10 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 24**.

Check whether the node can synchronize time properly with the NTP service on the active OMS node.

Step 24 Check whether the node can synchronize time with the NTP service on the active OMS node based on additional information of the alarm.

- If yes, go to **Step 25**.
- If no, go to **Step 35**.

Step 25 Check whether the synchronization with the NTP service on the active OMS node is faulty.

Log in to the alarm node as user **root** and run the **ntpq -np** command.

If an asterisk (*) exists before the IP address of the NTP service on the active OMS node in the command output, the synchronization is in normal state. The command output is as follows:

```
remote refid st t when poll reach delay offset jitter
==============================================================================
*10.10.10.162 .LOCL. 1 u 1 16 377 0.270 -1.562 0.014
```

If there is no asterisk (*) before the IP address of the NTP service on the active OMS node, as shown in the following command output, and the value of **refid** is **.INIT.**, the synchronization is abnormal.

```
remote refid st t when poll reach delay offset jitter
==============================================================================
10.10.10.162 .INIT. 1 u 1 16 377 0.270 -1.562 0.014
```

- If yes, go to **Step 26**.
- If no, go to **Step 38**.

Step 26 The NTP synchronization failure is typically caused by the system firewall. If the firewall can be disabled, run the **iptables -F** command to disable it. If the firewall cannot be disabled, run the **iptables -L** command to check the firewall configuration policy and ensure that the UDP port 123 is not disabled. (For details, see the firewall configuration policy of each system.)

Step 27 After 10 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 28**.

Step 28 Log in to the active OMS node as user **root** and run the following command to view the authentication key index field **1M**:

In SUSE Linux, run the **cat ${BIGDATA_HOME}/om-server/OMS/workspace/conf/ntp.keys** command.

In Red Hat Enterprise Linux or EulerOS, run the **cat ${BIGDATA_HOME}/om-server/OMS/workspace/conf/ntpkeys** command.

Step 29 Run the following command to check whether the key value is the same as that queried in **Step 28**:

In SUSE Linux, run the **diff ${BIGDATA_HOME}/om-server/OMS/workspace/conf/ntp.keys /etc/ntp.keys** command.

In Red Hat Enterprise Linux or EulerOS, run the **diff ${BIGDATA_HOME}/om-server/OMS/workspace/conf/ntpkeys /etc/ntp/ntpkeys** command.

NOTE

If the key values are the same, no result is returned after the command is executed. For example:

```
host01:~ # cat ${BIGDATA_HOME}/om-server/OMS/workspace/conf/ntp.keys
1 M sdYbq;o^CzEAWo<U=Tw5
host01:~ # diff ${BIGDATA_HOME}/om-server/OMS/workspace/conf/ntp.keys /etc/ntp.keys
host01:~ #
```

- If yes, go to **Step 30**.
- If no, go to **Step 38**.

Step 30 Run the **cat ${BIGDATA_HOME}/om-server/om/packaged-distributables/ntpKeyFile** command to check whether the key value is the same as that queried in **Step 28**. (Compare the key value with that of the authentication key index field **1M** queried in **Step 28**.)

- If yes, go to **Step 31**.
- If no, go to **Step 33**.

Step 31 Log in to the faulty node as user **root** and run the **cat /etc/ntp.keys** command in SUSE Linux (or the **cat /etc/ntp/ntpkeys** command in Red Hat Enterprise Linux) to check whether the key value is the same as the value queried in **Step 30** (use the key value of the authentication key index field **1M** for comparison).

- If yes, go to **Step 38**.
- If no, go to **Step 32**.

Step 32 Run the **su - omm** command to switch to user **omm**, change the key value of the authentication key index field **1M** in **${NODE_AGENT_HOME}/ntp.keys** (**${NODE_AGENT_HOME}/ntpkeys** in Red Hat Enterprise Linux) to the key value of **ntpKeyFile** in **Step 30**, and go to **Step 34**.

Step 33 Run the following commands as user **root** or **omm** to change the NTP key value of the active OMS node (change **ntp.keys** to **ntpkeys** in Red Hat Enterprise Linux):

cd ${BIGDATA_HOME}/om-server/OMS/workspace/conf

sed -i "`cat ntp.keys | grep -n '1 M'|awk -F ':' '{print $1}'`d" ntp.keys

echo "1 M `cat ${BIGDATA_HOME}/om-server/om/packaged-distributables/ntpKeyFile`" >>ntp.keys

Check whether the key value of the authentication key index field **1M** in **ntp.keys** is the same as that of **ntpKeyFile**.

- If yes, go to **Step 34**.
- If no, change the key value of the authentication key index field **1M** in **ntp.keys** to the key value of **ntpKeyFile** and go to **Step 34**.

Step 34 After 5 minutes, run the **service ntp restart** command to restart the NTP service on the active OMS node. After 15 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 38**.

Check whether the time deviation between the node and the NTP service on the active OMS node is large.

Step 35 Check whether the time deviation is large in additional information of the alarm.

- If yes, go to **Step 36**.
- If no, go to **Step 38**.

Step 36 On the **Hosts** tab page, select the host for which the alarm is generated, and choose **More** > **Stop All Instances** to stop all the services on the node.

If the time on the alarm node is later than that on the NTP service of the active OMS node, adjust the time of the alarm node. After adjusting the time, choose **More** > **Start All Instances** to start the services on the node.

If the time on the alarm node is earlier than that on the NTP service of the active OMS node, wait until the time deviation is due and adjust the time of the alarm node. After adjusting the time, choose **More** > **Start All Instances** to start the services on the node.

NOTE

If you do not wait, data loss may occur.

Step 37 After 10 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 38**.

Collect the fault information.

Step 38 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 39 In the **Services** area, select **NodeAgent** and **OmmServer**, and click **OK**. Expand the **Hosts** dialog box and select the alarm node and the active OMS node.

Step 40 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 41 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.9 ALM-12014 Partition Lost

Description

The system checks the partition status every 60 seconds. This alarm is generated when the system detects that a partition to which service directories are mounted is lost (because the device is removed or goes offline, or the partition is deleted). The system checks the partition status periodically.

This alarm must be manually cleared.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12014	Major	No

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
DirName	Specifies the directory for which the alarm is generated.
PartitionName	Specifies the device partition for which the alarm is generated.

Impact on the System

Service data fails to be written into the partition, and the service system runs abnormally.

Possible Causes

- The hard disk is removed.

- The hard disk is offline, or a bad sector exists on the hard disk.

Procedure

Step 1 On FusionInsight Manager, click **O&M** > **Alarm** > **Alarms**, and click ⌄ in the row where the alarm is located.

Step 2 Obtain **HostName**, **PartitionName** and **DirName** from **Location**.

Step 3 Check whether the disk of **PartitionName** on **HostName** is inserted to the correct server slot.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 Contact hardware engineers to remove the faulty disk.

Step 5 Log in to the **HostName** node where an alarm is reported and check whether there is a line containing **DirName** in the **/etc/fstab** file as user **root**.

- If yes, go to **Step 6**.
- If no, go to **Step 7**.

Step 6 Run the **vi /etc/fstab** command to edit the file and delete the line containing **DirName**.

Step 7 Contact hardware engineers to insert a new disk. For details, see the hardware product document of the relevant model. If the faulty disk is in a RAID group, configure the RAID group. For details, see the configuration methods of the relevant RAID controller card.

Step 8 Wait 20 to 30 minutes (The disk size determines the waiting time), and run the **mount** command to check whether the disk has been mounted to the **DirName** directory.

- If yes, manually clear the alarm. No further operation is required.
- If no, go to **Step 9**.

Collect fault information.

Step 9 On the FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 10 Select the **OmmServer** from the Services drop-down list and click **OK**.

Step 11 Set Start Date for log collection to 10 minutes ahead of the alarm generation time and End Date to 10 minutes behind the alarm generation time and click **Download**.

Step 12 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system does not automatically clear this alarm, and you need to manually clear the alarm.

Related Information

None

9.13.10 ALM-12015 Partition Filesystem Readonly

Description

The system checks the partition status every 60 seconds. This alarm is generated when the system detects that a partition to which service directories are mounted enters the read-only mode (due to a bad sector or a faulty file system). The system checks the partition status periodically.

This alarm is cleared when the system detects that the partition to which service directories are mounted exits from the read-only mode (because the file system is restored to read/write mode, the device is removed, or the device is formatted).

Attribute

Alarm ID	Alarm Severity	Auto Clear
12015	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
DirName	Specifies the directory for which the alarm is generated.
PartitionName	Specifies the device partition for which the alarm is generated.

Impact on the System

Service data fails to be written into the partition, and the service system runs abnormally.

Possible Causes

The hard disk is faulty, for example, a bad sector exists.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm > Alarms**, click ⌄ in the row where the alarm is located.

Step 2 Obtain **HostName** and **PartitionName** from **Location**. **HostName** is the node where the alarm is reported, and **PartitionName** is the partition of the faulty disk.

Step 3 Contact hardware engineers to check whether the disk is faulty. If the disk is faulty, remove it from the server.

Step 4 After the disk is removed, alarm **ALM-12014 Partition Lost** is reported. Handle the alarm. For details, see **ALM-12014 Partition Lost**. After the alarm **ALM-12014 Partition Lost** is cleared, alarm **ALM-12015 Partition Filesystem Readonly** is automatically cleared.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.11 ALM-12016 CPU Usage Exceeds the Threshold

Description

The system checks the CPU usage every 30 seconds and compares the actual CPU usage with the threshold. The CPU usage has a default threshold. This alarm is generated when the CPU usage exceeds the threshold for several times (configurable, 10 times by default) consecutively.

The alarm is cleared in the following two scenarios: The value of **Trigger Count** is 1 and the CPU usage is smaller than or equal to the threshold; the value of **Trigger Count** is greater than 1 and the CPU usage is smaller than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12016	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

Service processes respond slowly or become unavailable.

Possible Causes

- The alarm threshold or alarm smoothing times are incorrect.
- CPU configuration cannot meet service requirements. The CPU usage reaches the upper limit.

Procedure

Check whether the alarm threshold or alarm Trigger Count are correct.

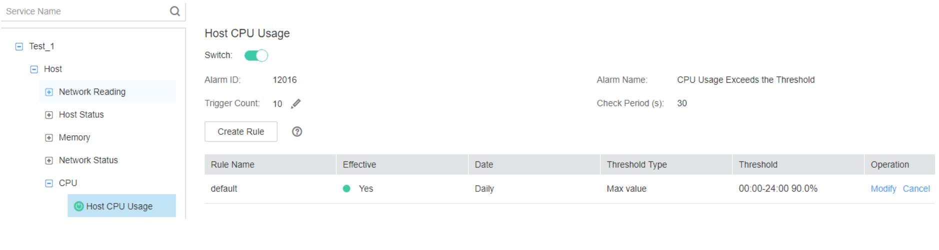
Step 1 Change the alarm threshold and alarm **Trigger Count** based on CPU usage.

On FusionInsight Manager, choose **O&M** > **Alarm** > **Thresholds >** *Name of the desired cluster* > **Host** > **CPU** > **Host CPU Usage** and change the alarm smoothing times based on CPU usage, as shown in **Figure 9-15**.

NOTE

This option defines the alarm check phase. **Trigger Count** indicates the alarm check threshold. An alarm is generated when the number of check times exceeds the threshold.

Figure 9-15 Setting alarm smoothing times



Host CPU Usage

Switch:

Alarm ID: 12016

Alarm Name: CPU Usage Exceeds the Threshold

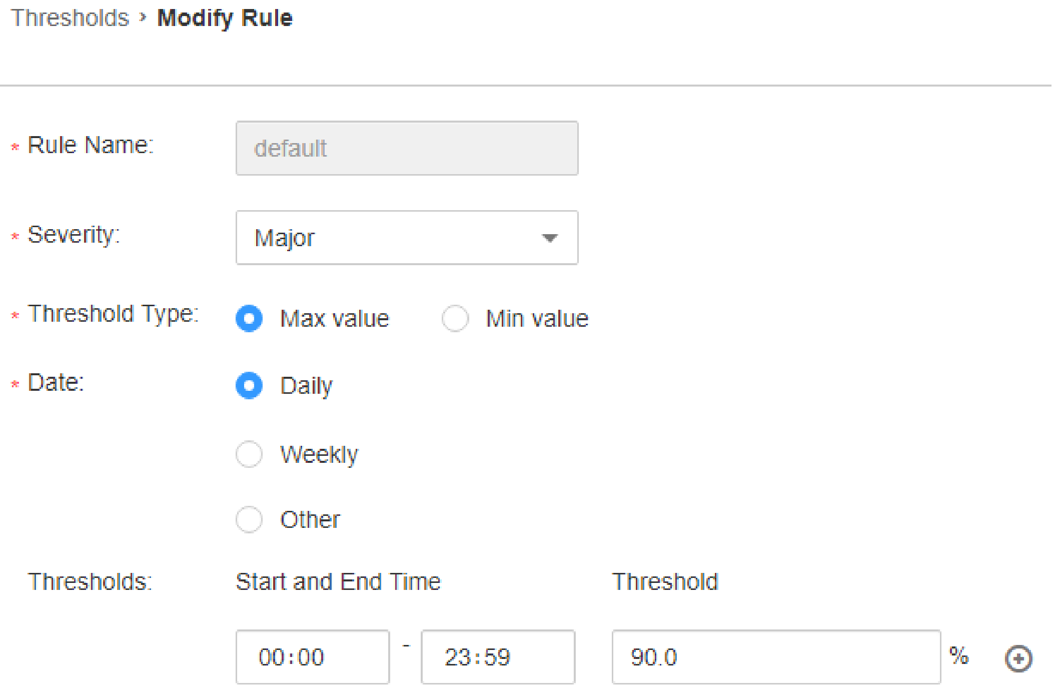
Trigger Count: 10

Check Period (s): 30

Rule Name	Effective	Date	Threshold Type	Threshold	Operation
default	● Yes	Daily	Max value	00:00-24:00 90.0%	Modify Cancel

On **Host CPU Usage** page and click **Modify** in the **Operation** column to change the alarm threshold, as shown in **Figure 9-16**.

Figure 9-16 Setting an alarm threshold



Thresholds › **Modify Rule**

* Rule Name: default

* Severity: Major

* Threshold Type: ● Max value ○ Min value

* Date: ● Daily

○ Weekly

○ Other

Thresholds:	Start and End Time	Threshold
	00:00 - 23:59	90.0 %

Step 2 After 2 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 3**.

Check whether the CPU usage reaches the upper limit.

Step 3 In the alarm list on FusionInsight Manager, click ∨ in the row where the alarm is located to view the alarm host address in the alarm details.

Step 4 On the **Hosts** page, click the node on which the alarm is reported.

Step 5 View the CPU usage for 5 minutes. If the CPU usage exceeds the threshold for multiple times, contact the system administrator to add more CPUs.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager in the active cluster, choose **O&M** > **Log > Download**.

Step 8 Select **OmmServer** from the **Service** and click **OK**.

Step 9 Set **Start Date** for log collection to 10 minutes ahead of the alarm generation time and **End Date** to 10 minutes behind the alarm generation time in **Time Range** and click **Download**.

Step 10 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.12 ALM-12017 Insufficient Disk Capacity

Description

The system checks the host disk usage of the system every 30 seconds and compares the actual disk usage with the threshold. The disk usage has a default threshold, this alarm is generated when the host disk usage exceeds the specified threshold.

When the **Trigger Count** is 1, this alarm is cleared when the usage of a host disk partition is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the usage of a host disk partition is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12017	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
PartitionName	Specifies the device partition for which the alarm is generated.

Name	Meaning
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

Service processes become unavailable.

Possible Causes

- The alarm threshold is incorrect.
- Disk configuration of the server cannot meet service requirements.

Procedure

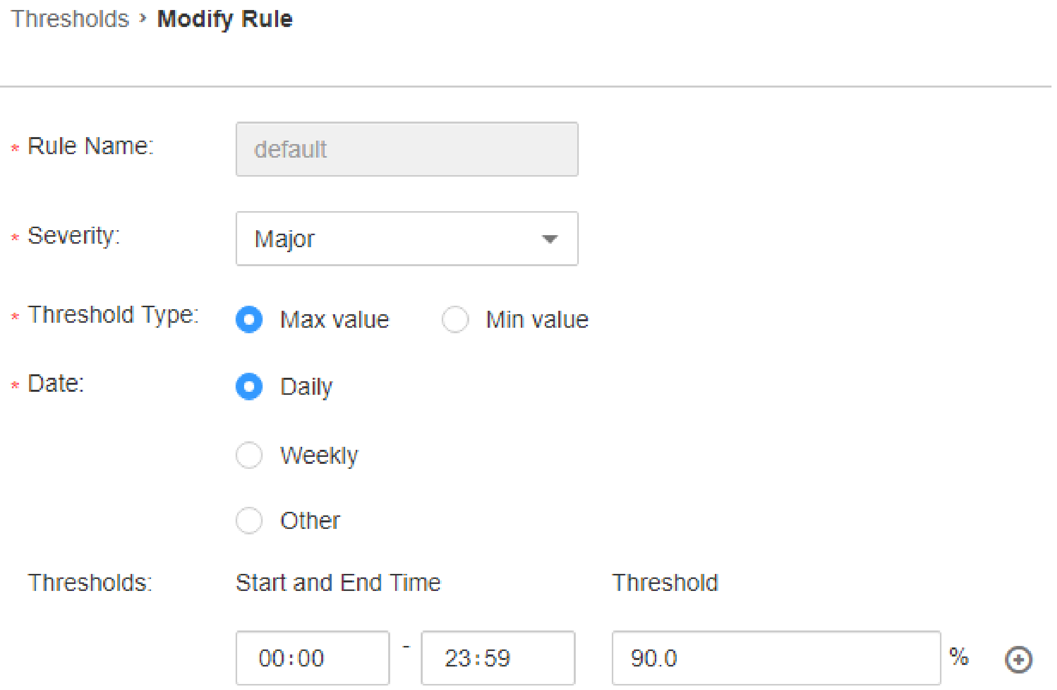
Check whether the alarm threshold is appropriate.

Step 1 Log in to FusionInsight Manager, choose **O&M** > **Alarm > Thresholds >** *Name of the desired cluster* > **Host** > **Disk** > **Disk Usage** and check whether the threshold (configurable, 90% by default) is appropriate.

- If yes, go to **Step 2**.
- If no, go to **Step 4**.

Step 2 Choose **O&M** > **Alarm > Thresholds >** *Name of the desired cluster* > **Host** > **Disk** > **Disk Usage** and click **Modify** in the **Operation** column to change the alarm threshold based on site requirements. As shown in **Figure 9-17**:

Figure 9-17 Setting an alarm threshold



Thresholds › Modify Rule

* Rule Name: default

* Severity: Major

* Threshold Type: ● Max value ○ Min value

* Date: ● Daily ○ Weekly ○ Other

Thresholds: Start and End Time 00:00 - 23:59 Threshold 90.0 %

Step 3 After 2 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check whether the disk usage reaches the upper limit.

Step 4 In the alarm list on FusionInsight Manager, click ⌄ in the row where the alarm is located to view the alarm host name and disk partition information in the alarm details.

Step 5 Log in to the node where the alarm is generated as user **root**.

Step 6 Run the **df -lmPT | awk '$2 != "iso9660"' | grep '^/dev/' | awk '{"readlink -m "$1 | getline real }{$1=real; print $0}' | sort -u -k 1,1** command to check the system disk partition usage. Check whether the disk is mounted to the following directories based on the disk partition name obtained in **Step 4**: **/**, **/opt**, **/tmp**, **/var**, **/var/log**, and **/srv/BigData**(can be customized).

- If yes, the disk is a system disk. Then go to **Step 10**.
- If no, the disk is not a system disk. Then go to **Step 7**.

Step 7 Run the **df -lmPT | awk '$2 != "iso9660"' | grep '^/dev/' | awk '{"readlink -m "$1 | getline real }{$1=real; print $0}' | sort -u -k 1,1** command to check the system disk partition usage. Determine the role of the disk based on the disk partition name obtained in **Step 4**.

Step 8 Check the disk service.

In MRS, check whether the disk service is HDFS, Yarn, Kafka, Supervisor.

- If yes, adjust the capacity. Then go to **Step 9**.
- If no, go to **Step 12**.

Step 9 After 2 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 12**.

Step 10 Run the **find / -xdev -size +500M -execls -l {} \;** command to check whether a file larger than 500 MB exists on the node and disk.

- If yes, go to **Step 11**.
- If no, go to **Step 12**.

Step 11 Handle the large file and check whether the alarm is cleared 2 minutes later.

- If yes, no further action is required.
- If no, go to **Step 12**.

Step 12 Contact the system administrator to expand the disk capacity.

Step 13 After 2 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 14**.

Collect fault information.

Step 14 On FusionInsight Manager, choose **O&M** > **Log > Download**.

Step 15 Select **OMS** from the **Service** and click **OK**.

Step 16 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 17 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.13 ALM-12018 Memory Usage Exceeds the Threshold

Description

The system checks the memory usage of the system every 30 seconds and compares the actual memory usage with the threshold. The memory usage has a default threshold, this alarm is generated when the value of the memory usage exceeds the threshold.

When the **Trigger Count** is 1, this alarm is cleared when the host memory usage is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the host memory usage is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12018	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.

Name	Meaning
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

Service processes respond slowly or become unavailable.

Possible Causes

- Memory configuration cannot meet service requirements. The memory usage reaches the upper limit.
- The SUSE 12.X OS has an earlier **free** command. The calculated memory usage cannot reflect the real-world memory usage.

Procedure

Perform the following operations if SUSE 12.X is used.

Step 1 Log in to any node in the cluster as user **root**, and run the **cat /etc/*-release** command to check whether the OS is SUSE 12.X as user **root**.

- If yes, go to **Step 2**.
- If no, go to **Step 4**.

Step 2 Run the **cat /proc/meminfo | grep Mem** command to check the real-world memory usage of the OS.

```
MemTotal: 263576192 kB
MemFree: 198283116 kB
MemAvailable: 227641452 kB
```

Step 3 Calculate the real-world memory usage: Memory usage = 1 - (Memory available/ Memory total)

- If the memory usage is lower than 90%, manually disable transferring from monitoring indicators to alarms.
- If the memory usage is higher than 90%, go to **Step 4**.

Expand the system.

Step 4 In the alarm list on FusionInsight Manager, click ⌄ in the row where the alarm is located to view the alarm host address in the alarm details.

Step 5 Log in to the host where the alarm is generated as user **root**.

Step 6 If the memory usage exceeds the threshold, perform memory capacity expansion.

Step 7 Run the command **free -m | grep Mem\: | awk '{printf("%s,", $3 * 100 / $2)}'** to check the system memory usage.

Step 8 Wait for 5 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect fault information.

Step 9 On the FusionInsight Manager in the active cluster, choose **O&M** > **Log > Download**.

Step 10 Select **OmmServer** from the **Servic**e and click **OK**.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.14 ALM-12027 Host PID Usage Exceeds the Threshold

Description

The system checks the PID usage every 30 seconds and compares the actual PID usage with the default PID usage threshold. This alarm is generated when the system detects that the PID usage exceeds the threshold.

When the **Trigger Count** is 1, this alarm is cleared when the PID usage is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the PID usage is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12027	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

No PID is available for new processes and service processes are unavailable.

Possible Causes

Too many processes are running on the node. You need to increase the value of **pid_max**.

Procedure

Increase the value of pid_max.

Step 1 In the alarm list on FusionInsight Manager, click ⌄ in the row where the alarm is located to view the alarm host address in the alarm details.

Step 2 Log in to the host where the alarm is generated as user **root**.

Step 3 Run the **cat /proc/sys/kernel/pid_max**command to check the value of **pid_max**.

Step 4 If the PID usage exceeds the threshold, run the command **echo** *new value* **> /proc/sys/kernel/pid_max** to enlarge the value of **pid_max**.

Example: **echo 65536 > /proc/sys/kernel/pid_max**

NOTE

The maximum value of **pid_max** is as follows:

- On 32-bit systems: 32768
- On 64-bit systems: 4194304 (2^22)

Step 5 Wait for 5 minutes, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager home page of the active cluster, choose **O&M** > **Log** > **Download**.

Step 7 Select all services from the **Service** and click **OK**.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.15 ALM-12028 Number of Processes in the D State on a Host Exceeds the Threshold

Description

The system checks the number of processes in the D state of user **omm** on the host every 30 seconds and compares the actual number with the threshold. The number of processes in the D state on the host has a default threshold range. This alarm is generated when the number of processes exceeds the threshold.

This alarm is cleared when the **Trigger Count** is **1** and the total number of processes in the D state of user **omm** on the host does not exceed the threshold. This alarm is cleared when the **Trigger Count** is greater than **1** and the total number of processes in the D state of user **omm** on the host is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12028	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.

Name	Meaning
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Excessive system resources are used and service processes respond slowly.

Possible Causes

The host responds slowly to I/O (disk I/O and network I/O) requests and some processes are in the D state and Z state.

Procedure

Check the processes in the D state.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and click ⌄ to view the IP address of the host for which the alarm is generated.

Step 2 Log in to the host for which the alarm is generated as user **root**. () Then run the **su - omm** command to switch to user **omm**.

Step 3 Run the following command as user **omm** to view the PID of the process that is in the D state:

```
ps -elf | grep -v "\[thread_checkio\]" | awk 'NR!=1 {print $2, $3, $4}' | grep omm | awk -F' ' '{print $1, $3}' | grep -E "Z|D" | awk '{print $2}'
```

Step 4 Check whether the command output is empty.

- If yes, the service process is running properly. Then go to **Step 6**.
- If no, go to **Step 5**.

Step 5 Switch to user **root** and run the **reboot** command to restart the host for which the alarm is generated. (Restarting a host is risky. Ensure that the service process is normal after the restart.)

Step 6 Check whether the alarm is cleared 5 minutes later.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect the fault information.

Step 7 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 8 Select **OMS** for **Service** and click **OK**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.16 ALM-12033 Slow Disk Fault

Description

- For HDDs, the alarm is triggered when any of the following conditions is met:
  - The system runs the **iostat** command every 3 seconds, and detects that the **svctm** value exceeds 1000 ms for 2 consecutive periods within 30 seconds.
  - The system runs the **iostat** command every 3 seconds, and detects that more than 50% of non-zero I/Os take more than 150 ms within 300s.
- For SSDs, the alarm is triggered when any of the following conditions is met:
  - The system runs the **iostat** command every 3 seconds, and detects that the **svctm** value exceeds 1000 ms for 10 consecutive periods within 30 seconds.
  - The system runs the **iostat** command every 3 seconds, and detects that more than 60% of I/Os take more than 20 ms within 300 seconds.

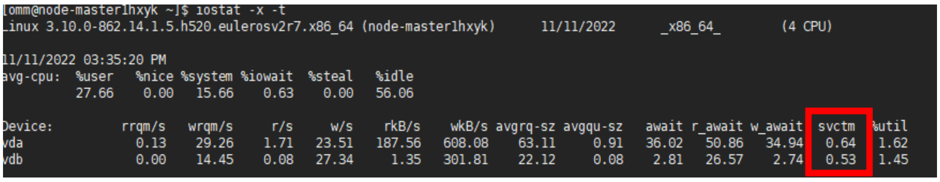
This alarm is automatically cleared when the preceding conditions have not been met for 15 minutes.

NOTE

The **svctm** value can be obtained as follows:

- MRS 3.1.0:

  Run the **iostat -x -t** command in the OS.



```
[omm@node-master1hxyk ~]$ iostat -x -t
Linux 3.10.0-862.14.1.5.h520.eulerosv2r7.x86_64 (node-master1hxyk)      11/11/2022      _x86_64_        (4 CPU)

11/11/2022 03:35:20 PM
avg-cpu:  %user   %nice %system %iowait  %steal   %idle
          27.66    0.00   15.66    0.63    0.00   56.06

Device:         rrqm/s   wrqm/s     r/s     w/s    rkB/s    wkB/s avgrq-sz avgqu-sz   await r_await w_await  svctm  %util
vda               0.13    29.26    1.71   23.51   187.56   608.08    63.11     0.91   36.02   50.86   34.94   0.64   1.62
vdb               0.00    14.45    0.08   27.34     1.35   301.81    22.12     0.08    2.81   26.57    2.74   0.53   1.45
```

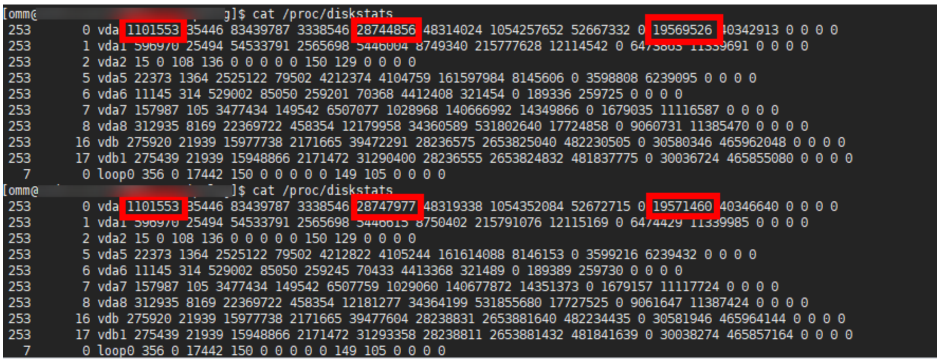
- Versions later than MRS 3.1.0:

svctm = (tot_ticks_new - tot_ticks_old)/(rd_ios_new + wr_ios_new - rd_ios_old - wr_ios_old)

If **rd_ios_new + wr_ios_new - rd_ios_old - wr_ios_old** is **0**, then **svctm** is **0**.

The parameters can be obtained as follows:

The system runs the **cat /proc/diskstats** command every 3 seconds to collect data. For example:



```
[omm@[redacted]]$ cat /proc/diskstats
 253       0 vda 1101553 35446 83439787 3338546 28744856 48314024 1054257652 52667332 0 19569526 40342913 0 0 0 0
 253       1 vda1 596970 25494 54533791 2565698 5446004 8749340 215777628 12114542 0 6473683 11339691 0 0 0 0
 253       2 vda2 15 0 108 136 0 0 0 0 0 150 129 0 0 0 0
 253       5 vda5 22373 1364 2525122 79502 4212374 4104759 161597984 8145606 0 3598808 6239095 0 0 0 0
 253       6 vda6 11145 314 529002 85050 259201 70368 4412408 321454 0 189336 259725 0 0 0 0
 253       7 vda7 157987 105 3477434 149542 6507077 1028968 140666992 14349866 0 1679035 11116587 0 0 0 0
 253       8 vda8 312935 8169 22369722 458354 12179958 34360589 531802640 17724858 0 9060731 11385470 0 0 0 0
 253      16 vdb 275920 21939 15977738 2171665 39472291 28236575 2653825040 482230505 0 30580346 465962048 0 0 0 0
 253      17 vdb1 275439 21939 15948866 2171472 31290400 28236555 2653824832 481837775 0 30036724 465855080 0 0 0 0
   7       0 loop0 356 0 17442 150 0 0 0 0 0 149 105 0 0 0 0
[omm@[redacted]]$ cat /proc/diskstats
 253       0 vda 1101553 35446 83439787 3338546 28747977 48319338 1054352084 52672715 0 19571460 40346640 0 0 0 0
 253       1 vda1 596970 25494 54533791 2565698 5446615 8750402 215791076 12115169 0 6474429 11339985 0 0 0 0
 253       2 vda2 15 0 108 136 0 0 0 0 0 150 129 0 0 0 0
 253       5 vda5 22373 1364 2525122 79502 4212822 4105244 161614088 8146153 0 3599216 6239432 0 0 0 0
 253       6 vda6 11145 314 529002 85050 259245 70433 4413368 321489 0 189389 259730 0 0 0 0
 253       7 vda7 157987 105 3477434 149542 6507759 1029060 140677872 14351373 0 1679157 11117724 0 0 0 0
 253       8 vda8 312935 8169 22369722 458354 12181277 34364199 531855680 17727525 0 9061647 11387424 0 0 0 0
 253      16 vdb 275920 21939 15977738 2171665 39477604 28238831 2653881640 482234435 0 30581946 465964144 0 0 0 0
 253      17 vdb1 275439 21939 15948866 2171472 31293358 28238811 2653881432 481841639 0 30038274 465857164 0 0 0 0
   7       0 loop0 356 0 17442 150 0 0 0 0 0 149 105 0 0 0 0
```

In these two commands:

In the data collected for the first time, the number in the fourth column is the **rd_ios_old** value, the number in the eighth column is the **wr_ios_old** value, and the number in the thirteenth column is the **tot_ticks_old** value.

In the data collected for the second time, the number in the fourth column is the **rd_ios_new** value, the number in the eighth column is the **wr_ios_new** value, and the number in the thirteenth column is the **tot_ticks_new** value.

In this case, the value of **svctm** is as follows:

(19571460 - 19569526)/(1101553 + 28747977 - 1101553 - 28744856) = 0.6197

Attribute

Alarm ID	Alarm Severity	Auto Clear
12033	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
DiskName	Specifies the disk for which the alarm is generated.

Impact on the System

Service performance deteriorates, service processing capabilities become poor, and services may be unavailable.

Possible Causes

The disk is aged or has bad sectors.

Procedure

Check the disk status.

Step 1 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Alarm** > **Alarms**.

Step 2 View the detailed information about the alarm. Check the values of **HostName** and **DiskName** in the location information to obtain the information about the faulty disk for which the alarm is generated.

Step 3 Check whether the node for which the alarm is generated is in a virtualization environment.

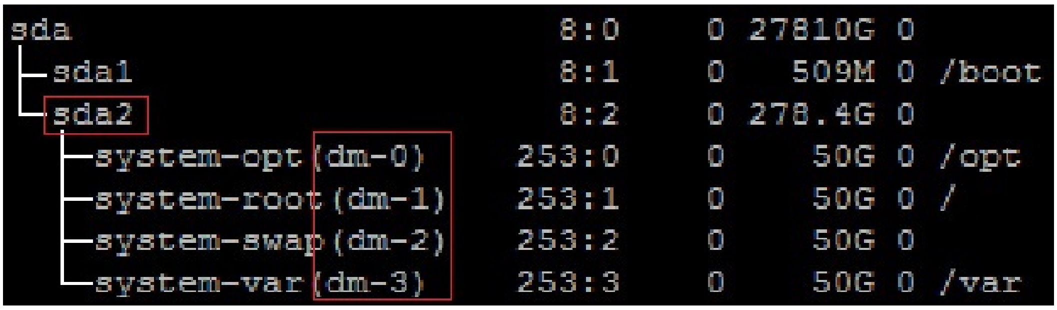
- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 4 Check whether the storage performance provided by the virtualization environment meets the hardware requirements. Then, go to **Step 5**.

Step 5 Log in to the alarm node as user **root**, run the **df -h** command, and check whether the command output contains the value of the **DiskName** field.

- If yes, go to **Step 7**.
- If no, go to **Step 6**.

Step 6 Run the **lsblk** command to check whether the mapping between the value of **DiskName** and the disk has been created.



```
sda                          8:0     0 27810G 0
├─sda1                       8:1     0   509M 0 /boot
└─sda2                       8:2     0 278.4G 0
  ├─system-opt (dm-0)      253:0     0    50G 0 /opt
  ├─system-root (dm-1)     253:1     0    50G 0 /
  ├─system-swap (dm-2)     253:2     0    50G 0
  └─system-var (dm-3)      253:3     0    50G 0 /var
```

- If yes, go to **Step 7**. .
- If no, go to **Step 22**.

Step 7 Log in to the alarm node as user **root**, run the **lsscsi | grep "/dev/sd[x]"** command to view the disk information, and check whether RAID has been set up.

NOTE

In the command, **/dev/sd[x]** indicates the disk name obtained in **Step 2**.

Example:

lsscsi | grep "/dev/sda"

In the command output, if **ATA**, **SATA**, or **SAS** is displayed in the third line, the disk has not been organized into a RAID group. If other information is displayed, RAID has been set up.

- If yes, go to **Step 12**.
- If no, go to **Step 8**.

Step 8 Run the **smartctl -i /dev/sd[x]** command to check whether the hardware supports the SMART tool.

Example:

smartctl -i /dev/sda

In the command output, if "SMART support is: Enabled" is displayed, the hardware supports SMART. If "Device does not support SMART" or other information is displayed, the hardware does not support SMART.

- If yes, go to **Step 9**.
- If no, go to **Step 17**.

Step 9 Run the **smartctl -H --all /dev/sd[x]** command to check basic SMART information and determine whether the disk is working properly.

Example:

smartctl -H --all /dev/sda

Check the value of **SMART overall-health self-assessment test result** in the command output. If the value is **FAILED**, the disk is faulty and needs to be replaced. If the value is **PASSED**, check the value of **Reallocated_Sector_Ct** or **Elements in grown defect list**. If the value is greater than 100, the disk is faulty and needs to be replaced.

- If yes, go to **Step 10**.

- If no, go to **Step 18**.

Step 10 Run the **smartctl -l error -H /dev/sd[x]** command to check the Glist of the disk and determine whether the disk is normal.

Example:

smartctl -l error -H /dev/sda

Check the **Command/Feature_name** column in the command output. If **READ SECTOR(S)** or **WRITE SECTOR(S)** is displayed, the disk has bad sectors. If other errors occur, the disk circuit board is faulty. Both errors indicate that the disk is abnormal and needs to be replaced.

If "No Errors Logged" is displayed, no error log exists. You can trigger the disk SMART self-check.

- If yes, go to **Step 11**.
- If no, go to **Step 18**.

Step 11 Run the **smartctl -t long /dev/sd[x]** command to trigger the disk SMART self-check. After the command is executed, the time when the self-check is to be completed is displayed. After the self-check is completed, repeat **Step 9** and **Step 10** to check whether the disk is working properly.

Example:

smartctl -t long /dev/sda

- If yes, go to **Step 17**.
- If no, go to **Step 18**.

Step 12 Run the **smartctl -d [sat|scsi]+megaraid,[DID] -H --all /dev/sd[x]** command to check whether the hardware supports SMART.

NOTE

- In the command, **[sat|scsi]** indicates the disk type. Both types need to be used.
- **[DID]** indicates the slot information. Slots 0 to 15 need to be used.

For example, run the following commands in sequence:

smartctl -d sat+megaraid,0 -H --all /dev/sda

smartctl -d sat+megaraid,1 -H --all /dev/sda

smartctl -d sat+megaraid,2 -H --all /dev/sda

...

Try the command combinations of different disk types and slot information. If "SMART support is: Enabled" is displayed in the command output, the disk supports SMART. Record the parameters of the disk type and slot information when a command is successfully executed. If "SMART support is: Enabled" is not displayed in the command output, the disk does not support SMART.

- If yes, go to **Step 13**.
- If no, go to **Step 16**.

Step 13 Run the **smartctl -d [sat|scsi]+megaraid,[DID] -H --all /dev/sd[x]** command recorded in **Step 12** to check basic SMART information and determine whether the disk is normal.

Example:

smartctl -d sat+megaraid,2 -H --all /dev/sda

Check the value of **SMART overall-health self-assessment test result** in the command output. If the value is **FAILED**, the disk is faulty and needs to be replaced. If the value is **PASSED**, check the value of **Reallocated_Sector_Ct** or **Elements in grown defect list**. If the value is greater than 100, the disk is faulty and needs to be replaced.

- If yes, go to **Step 14**.
- If no, go to **Step 18**.

Step 14 Run the **smartctl -d [sat|scsi]+megaraid,[DID] -l error -H /dev/sd[x]** command to check the Glist of the disk and determine whether the hard disk is working properly.

Example:

smartctl -d sat+megaraid,2 -l error -H /dev/sda

Check the **Command/Feature_name** column in the command output. If **READ SECTOR(S)** or **WRITE SECTOR(S)** is displayed, the disk has bad sectors. If other errors occur, the disk circuit board is faulty. Both errors indicate that the disk is abnormal and needs to be replaced.

If "No Errors Logged" is displayed, no error log exists. You can trigger the disk SMART self-check.

- If yes, go to **Step 15**.
- If no, go to **Step 18**.

Step 15 Run the **smartctl -d [sat|scsi]+megaraid,[DID] -t long /dev/sd[x]** command to trigger the disk SMART self-check. After the command is executed, the time when the self-check is to be completed is displayed. After the self-check is completed, repeat **Step 13** and **Step 14** to check whether the disk is working properly.

Example:

smartctl -d sat+megaraid,2 -t long /dev/sda

- If yes, go to **Step 17**.
- If no, go to **Step 18**.

Step 16 If the configured RAID controller card does not support SMART, the disk does not support SMART. In this case, use the check tool provided by the corresponding RAID controller card vendor to rectify the fault. Then go to **Step 17**.

For example, LSI is a MegaCLI tool.

Step 17 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**, click **Clear** in the **Operation** column of the alarm, and check whether the alarm is reported on the same disk again.

If the alarm is reported for three times, replace the disk.

- If yes, go to **Step 18**.
- If no, no further action is required.

Replace the disk.

Step 18 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Alarm** > **Alarms**.

Step 19 View the detailed information about the alarm. Check the values of **HostName** and **DiskName** in the location information to obtain the information about the faulty disk for which the alarm is reported.

Step 20 Replace the disk.

Step 21 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 22**.

Collect the fault information.

Step 22 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 23 Select **OMS** for **Service** and click **OK**.

Step 24 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 25 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.17 ALM-12034 Periodical Backup Failure

Description

The system executes the periodic backup task every 60 minutes. This alarm is generated when a periodical backup task fails to be executed. This alarm is cleared when the next backup task is executed successfully.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12034	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
TaskName	Specifies the task.

Impact on the System

There are not available backup packages for a long time, so the system cannot be restored in case of exceptions.

Possible Causes

The alarm cause depends on the task details. Handle the alarm according to the logs and alarm details.

Procedure

Check whether the disk space is sufficient.

Step 1 In the FusionInsight Manager portal, click **O&M > Alarm > Alarms**.

Step 2 In the alarm list, click ⌄ in the row where the alarm is located and obtain **TaskName** from **Location**.

Step 3 Choose **O&M** > **Backup and Restoration > Backup Management**.

Step 4 Search for the backup task based on **TaskName** and click **More** in the **Operation** column. In the displayed dialog box, click **View History** and view the task details.

Step 5 In the displayed dialog box and click ⌄ to check whether the following message is displayed: Failed to backup xx due to insufficient disk space, move the data in the xx directory to other directories.

- If yes, go to **Step 6**.
- If no, go to **Step 13**.

Step 6 Choose **Backup Path** > **View** and obtain the **Backup Path**.

Step 7 Log in to the node as user **root** and run the following command to check the node mounting details:

```
df -h
```

Step 8 Check whether the available space of the node to which the backup path is mounted is less than 20 GB.

- If yes, go to **9**.
- If no, go to **Step 13**.

Step 9 Check whether there are many backup packages in the backup directory.

- If yes, go to **Step 10**.
- If no, go to **Step 13**.

Step 10 Enable the available space of the node to which the backup directory is mounted to be greater than 20 GB by moving backup packages out of the backup directory or delete the backup packages.

Step 11 After the problem is resolved, perform the backup task again and check whether the backup task execution is successful.

- If yes, go to **Step 12**.
- If no, go to **Step 13**.

Step 12 After 2 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 13**.

Collect fault information.

Step 13 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 14 Select **Controller** from the **Service** and click **OK**.

Step 15 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 16 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.18 ALM-12035 Unknown Data Status After Recovery Task Failure

Description

After the recovery task fails, the system automatically rolls back every 60 minutes. If the rollback fails, data may be lost. If this occurs, an alarm is reported. This alarm is cleared when the next recovery task execution is successful.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12035	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
TaskName	Specifies the task.

Impact on the System

After the recovery task fails, the system automatically rolls back. If the rollback fails, data may be lost or the data status may be unknown, which may affect services.

Possible Causes

The alarm cause depends on the task details. Handle the alarm according to the logs and alarm details.

Procedure

Collect fault information.

Step 1 In the FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services**, and check whether the running status of the component meets the requirements. (The OMS and DBService must be in the normal state, and other components must be stopped.)

- If yes, go to **Step 9**.
- If no, go to **Step 2**.

Step 2 Restore the component status as required and start the recovery task again.

Step 3 Log in to the FusionInsight Manager portal and click **O&M** > **Alarm** > **Alarms**.

Step 4 In the alarm list, click ⌄ in the row where the alarm is located to obtain **TaskName** from **Location**.

Step 5 Choose **O&M** > **Backup and Restoration > Restoration Management**.

Step 6 Find the restoration task by **Task Name** and view the task details.

Step 7 Perform the recovery task again and check whether the recovery task execution is successful.

- If yes, go to **8**.
- If no, go to **9**.

Step 8 After 2 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **9**.

Collect fault information.

Step 9 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 10 Select **Controller** from the **Service** and click **OK**.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.19 ALM-12037 NTP Server Abnormal

Description

The system checks the NTP server status every 60 seconds. This alarm is generated when the system detects that the NTP server is abnormal for 10 consecutive times.

This alarm is cleared when the NTP server recovers.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12037	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the IP address of the NTP server for which the alarm is generated.

Impact on the System

The NTP server configured on the active OMS node is abnormal. In this case, the active OMS node cannot synchronize time with the NTP server and a time offset may be generated in the cluster.

Possible Causes

- The NTP server network is abnormal.
- The NTP server authentication fails.
- The NTP server time cannot be obtained.
- The time obtained from the NTP server is not continuously updated.

Procedure

Check the NTP server network.

Step 1 On the FusionInsight Manager portal, click **O&M** > **Alarm** > **Alarms** and click ⌄ in the row where the alarm is located.

Step 2 View the alarm additional information to check whether the NTP server fails to be pinged.

- If yes, go to **Step 3**.
- If no, go to **Step 4**.

Step 3 Contact the network administrator to check the network configuration and ensure that the network between the NTP server and the active OMS node is normal. Then, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check whether the NTP server authentication fails.

Step 4 Log in to the active OMS node as user **root**.

Step 5 Run the following command to check the status of the resources on the active and standby nodes:

```
su - omm
sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh
```

- If "chrony" is displayed in the **ResName** column of the command output, go to **Step 6**.
- If "ntp" is displayed in the **ResName** column, go to **Step 7**.

NOTE

If both "chrony" and "ntp" are displayed in the **ResName** column of the command output, the NTP service mode is being switched. Wait for 10 minutes and perform **Step 5** again. If both "chrony" and "ntp" still exist in the **ResName** column, contact O&M personnel.

Step 6 Run the command **chronyc sources** to check whether the NTP server authentication fails.

If the value of **Reach** for chrony is **0**, the connection or authentication fails.

- If yes, go to **Step 12**.
- If no, go to **Step 8**.

Step 7 Run the command **ntpq -np** to check whether the NTP server authentication fails.

If **refid** of the NTP server is **.AUTH.**, the authentication fails.

- If yes, go to **Step 12**.
- If no, go to **Step 8**.

Check whether the time can be obtained from the NTP server.

Step 8 View the alarm additional information to check whether the time can be obtained from the NTP server.

- If yes, go to **Step 9**.
- If no, go to **Step 10**.

Step 9 Contact the provider of the NTP server to rectify the NTP server fault. After the NTP server is normal, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 10**.

Check whether the time obtained from the NTP server is not continuously updated.

Step 10 View the alarm additional information to check whether the time obtained from the NTP server is not continuously updated.

- If yes, go to **Step 11**.
- If no, go to **Step 12**.

Step 11 Contact the provider of the NTP server to rectify the NTP server fault. After the NTP server is normal, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 12**.

Collect fault information.

Step 12 On the FusionInsight Manager, choose **O&M** > **Log > Download**.

Step 13 Select **NodeAgent** and **OmmServer** from the **Service** and click **OK**.

Step 14 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 15 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.20 ALM-12038 Monitoring Indicator Dumping Failure

Description

After monitoring indicator dumping is configured on FusionInsight Manager, the system checks the monitoring indicator dumping result at the dumping interval (60 seconds by default). This alarm is generated when the dumping fails.

This alarm is cleared when dumping is successful.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12038	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The upper-layer management system cannot obtain monitoring indicators from the FusionInsight Manager system.

Possible Causes

- The server cannot be connected.
- The save path on the server cannot be accessed.
- The monitoring indicator file fails to be uploaded.

Procedure

Check whether the server connection is normal.

Step 1 Check whether the network between the FusionInsight Manager system and the server is normal.

- If yes, go to **Step 3**.
- If no, go to **Step 2**.

Step 2 Contact the network administrator to recover the network and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 3**.

Step 3 Choose **System** > **Interconnection > Upload Performance Data** and check whether the FTP username, password, port, dump mode, and public key configured on the upload performance data page are consistent with the configuration on the server.

- If yes, go to **Step 5**.
- If no, go to **Step 4**.

Step 4 Enter the correct configuration information, click **OK**, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check the permission of the save path on the server is correct.

Step 5 Choose **System** > **Interconnection > Upload Performance Data** and check the configuration items **FTP Username**, **Save Path**, and **Dump Mode**.

- If the dump mode is FTP, go to **Step 6**.
- If the dump mode is SFTP, go to **Step 7**.

Step 6 Log in to the server in FTP mode. In the default path, check whether **FTP Username** has the read and write permission of the relative path **Save Path**.

- If yes, go to **Step 9**.
- If no, go to **Step 8**.

Step 7 Log in to the server in SFTP mode and check whether **FTP Username** has the read and write permission of the absolute path **Save Path**.

- If yes, go to **Step 9**.
- If no, go to **Step 8**.

Step 8 Add the read and write permission and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Check whether the save path on the server has sufficient disk space.

Step 9 Log in to the server and check whether the save path has sufficient disk space.

- If yes, go to **Step 11**.
- If no, go to **Step 10**.

Step 10 Delete unnecessary files or go to the monitoring indicator dumping configuration page to change the save path. Then, check whether the save path has sufficient disk space.

- If yes, no further action is required.
- If no, go to **Step 11**.

Collect fault information.

Step 11 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 12 Select **OMS** from the **Service** and click **OK**.

Step 13 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 14 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.21 ALM-12039 Active/Standby OMS Databases Not Synchronized

Description

The system checks the data synchronization status between the active and standby OMS Databases every 10 seconds. This alarm is generated when the synchronization status cannot be queried for 30 consecutive times or when the synchronization status is abnormal.

This alarm is cleared when the data synchronization status becomes normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12039	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Local GaussDB HA IP	Specifies the HA IP address of the local GaussDB.
Peer GaussDB HA IP	Specifies the HA IP address of the peer GaussDB.
SYNC_PERCENT	Specifies the synchronization percentage.

Impact on the System

When data is not synchronized between the active and standby OMS Databases, data may be lost or abnormal if the active instance becomes abnormal.

Possible Causes

- The network between the active and standby nodes is unstable.
- The standby OMS Database is abnormal.
- The standby node disk space is full.

Procedure

Check whether the network between the active and standby nodes is normal.

Step 1 Log in to FusionInsight Manager, click **O&M > Alarm > Alarms**, click ∨ in the row where the alarm is located, and query the standby OMS Database IP address.

Step 2 Log in to the active OMS Database node as user **root**.

Step 3 Run the **ping** *Standby OMS Database heartbeat IP address* command to check whether the standby OMS Database node is reachable.

- If yes, go to **Step 6**.
- If no, go to **Step 4**.

Step 4 Contact the network administrator to check whether the network is faulty.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 Rectify the network fault and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Check whether the standby OMS Database is normal.

Step 6 Log in to the standby OMS Database node as user **root**.

Step 7 Run the **su - omm** command to switch to user **omm**.

Step 8 Go to the **${BIGDATA_HOME}/om-server/om/sbin/** directory and run the **./status-oms.sh** command to check whether the OMS Database resource status of the standby DBService is normal. In the command output, check whether the following information is displayed in the row where **ResName** is **gaussDB**:

For example:

```
10_10_10_231 gaussDB Standby_normal Normal Active_standby
```

- If yes, go to **Step 9**.
- If no, go to **Step 16**.

Check whether the standby node disk space is full.

Step 9 Log in to the standby OMS Database node as user **root**.

Step 10 Run the **su - omm** command to switch to user **omm**.

Step 11 Run the **echo ${BIGDATA_DATA_HOME}/dbdata_om** command to obtain the OMS Database data directory.

Step 12 Run the **df -h** command to view the system disk partition usage information.

Step 13 Check whether the disk where the OMS Database data directory is mounted is full.

- If yes, go to **Step 14**.
- If no, go to **Step 16**.

Step 14 Expand the disk capacity.

Step 15 After the disk capacity is expanded, wait 2 minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 16**.

Collect fault information.

Step 16 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 17 Select **OMMServer** from the **Service** and click **OK**.

Step 18 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 19 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.22 ALM-12040 Insufficient System Entropy

Description

The system checks the entropy for five consecutive times at 00:00 every day. Specifically, the system checks whether rng-tools or haveged has been enabled and correctly configured. If neither is configured, the system continues to check the entropy. If the entropy is less than 100 for five consecutive times, this alarm is reported.

This alarm is cleared when the system detects that the true random number mode has been configured, the random number parameters have been configured in the pseudo-random number mode, or neither mode is configured but the entropy of the OS is greater than or equal to 100 in at least one of five entropy checks.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12040	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.

Name	Meaning
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The system is not running properly.

Possible Causes

- rng-tools or haveged has not been installed or started.
- The entropy of the OS is smaller than 100 for multiple consecutive times.

Procedure

Check whether haveged or rng-tools has been installed or started.

Step 1 Log in to FusionInsight Manager and choose **O&M** > **Alarm** > **Alarms**.

Step 2 Check the value of **HostName** in the **Location** area to obtain the name of the host for which the alarm is generated.

Step 3 Log in to the node for which the alarm is generated as user **root**.

Step 4 Run the **/bin/rpm -qa | grep -w "haveged"** command to check the haveged installation status and check whether the command output is empty.

- If yes, go to **Step 6**.
- If no, go to **Step 5**.

Step 5 Run the **/sbin/service haveged status |grep "running"** command and check the command output.

- If the command is executed successfully, haveged has been installed and configured correctly and is running properly. Go to **Step 8**.
- If the command fails to execute, haveged is not running properly. Run the following command to manually restart haveged and go to **Step 9**:

  systemctl restart haveged.service

Step 6 Run the **/bin/rpm -qa | grep -w "rng-tools"** command to check the rng-tools installation and check whether the command output is empty.

- If yes, contact the OS vendor to install and start haveged or rng-tools. Then go to **Step 9**.
- If no, go to **Step 7**.

Step 7 Run the **ps -ef | grep -v "grep" | grep rngd | tr -d " " | grep "\-r/dev/urandom"** command and check the command output.

- If the command is executed successfully, rngd has been installed and configured correctly and is running properly. Go to **Step 8**.

- If the command fails to execute, rngd is not running properly. Run the following command to manually restart rngd and go to **Step 9**:

  systemctl restart rngd.service

Check the entropy of the OS.

Step 8 Manually check the entropy of the OS.

Log in to the target node as user **root** and run the **cat /proc/sys/kernel/random/entropy_avail** command to check whether the entropy of the OS meets cluster installation requirements (no less than 100).

- If yes, the entropy of the OS is not less than 100. Go to **Step 9**.
- If no, the entropy of the OS is less than 100. Use either of the following methods and go to **Step 9**.
  - Method 1: Use haveged (true random number mode). Contact the OS vendor to install and start haveged.

    In Kylin, run the following command:

    vi /usr/lib/systemd/system/haveged.service

    Configure **Type**, **ExecStar**, **SuccessExitStatus**, and **Restart** in **[Service]** as follows:

    ```
    Type=simple
    ExecStar=/usr/sbin/haveged -w 1024 -v 1 –Foreground
    SuccessExitStatus=137 143
    Restart=always
    ```
  - Method 2: Use rng-tools (pseudo-random number mode). Contact the OS vendor to install and start rng-tools and configure it based on the OS type.
    - In Red Hat Linux or CentOS, run the following commands:

      echo 'EXTRAOPTIONS="-r /dev/urandom -o /dev/random -t 1 -i"' >> /etc/sysconfig/rngd

      service rngd start

      chkconfig rngd on
    - In SUSE, run the following commands:

      rngd -r /dev/urandom -o /dev/random

      echo "rngd -r /dev/urandom -o /dev/random" >> /etc/rc.d/after.local
    - In Kylin, run the following command as user **root** on the node where the alarm is reported:

      vi /usr/lib/systemd/system/rngd.service

      Change the value of **ExecStart** in **[Service]** as follows:

      ```
      ExecStart=/sbin/rngd -f -r /dev/urandom -s 2048
      ```

Step 9 Wait until the system to check the entropy at 00:00 on the following day and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 10**.

Collect fault information.

Step 10 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 11 Select **NodeAgent** for **Service** and click **OK**.

Step 12 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 13 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

9.13.23 ALM-12041 Incorrect Permission on Key Files

Description

The system checks whether the permission, user, and user group information about critical directories or files is normal every 5 minutes. This alarm is generated when the information is abnormal.

This alarm is cleared when the information becomes normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12041	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
PathName	Specifies the path or name of the abnormal file.

Impact on the System

System functions are unavailable.

Possible Causes

The file permission is abnormal or the file is lost due to a user manually modified information such as the file permission, user, and user group, or the system is powered off unexpectedly.

Procedure

Check whether the abnormal file exists and whether the permission on the abnormal file is correct.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms**.

Step 2 Check the value of **HostName** to obtain the host name involved in this alarm. Check the value of **PathName** to obtain the path or name of the abnormal file.

Step 3 Log in to the node for which the alarm is generated as user **root**.

Step 4 Run the **ll** *pathName* command, where *pathName* indicates the name of the abnormal file to obtain the user, permission, and user group information about the file or directory.

Step 5 Go to **${BIGDATA_HOME}/om-agent/nodeagent/etc/agent/autocheck** directory. Then run the **vi keyfile** command and search for the name of the abnormal file and check the due permission of the file.

NOTE

To ensure proper configuration synchronization between the active and standby OMS servers, files, directories, and files and sub-directories in the directories configured in **$OMS_RUN_PATH/workspace/ha/module/hasync/plugin/conf/filesync.xml** will also be monitored except files and directories in **keyfile**. User **omm** must have read and write permissions of files and read and execute permissions of directories.

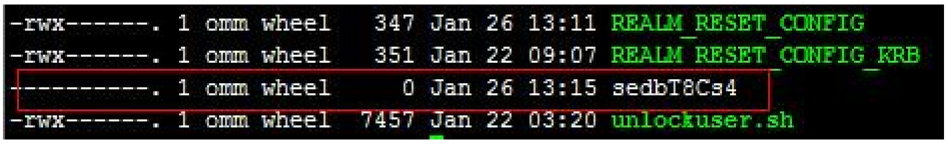
Step 6 Compare the real-world permission of the file with the due permission obtained in **Step 5** and correct the permission, user, and user group information for the file.

Step 7 Wait a hour and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

NOTE

If the disk partition where the cluster installation directory resides is used up, some temporary files will be generated in the program installation directory when running the **sed** command fails. Users do not have the read, write, and execute permissions of these temporary files. The system reports an alarm indicating that permissions of temporary files are abnormal if these files are within the monitoring range of the alarm. Perform the preceding alarm handling processes to clear the alarm. Alternatively, you can directly delete the temporary files after confirming that files with abnormal permissions are temporary. The temporary file generated after a **sed** command execution failure is similar to the following.



```
-rwx------. 1 omm wheel   347 Jan 26 13:11 REALM_RESET_CONFIG
-rwx------. 1 omm wheel   351 Jan 22 09:07 REALM_RESET_CONFIG_KRB
----------. 1 omm wheel     0 Jan 26 13:15 sedbT8Cs4
-rwx------. 1 omm wheel  7457 Jan 22 03:20 unlockuser.sh
```

Collect fault information.

Step 8 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 9 Select **NodeAgent** from the **Service** and click **OK**.

Step 10 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.24 ALM-12042 Incorrect Configuration of Key Files

Description

The system checks whether critical configurations are correct every 5 minutes. This alarm is generated when the configurations are abnormal.

This alarm is cleared when the configurations become normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12042	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.

Name	Meaning
PathName	Specifies the path or name of the abnormal file.

Impact on the System

Functions related to the file are abnormal.

Possible Causes

The file configuration is modified manually or the system is powered off unexpectedly.

Procedure

Check abnormal file configuration.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms**.

Step 2 Check the value of **HostName** to obtain the host name involved in this alarm. Check the value of **PathName** to obtain the path or name of the abnormal file.

Step 3 Log in to the node for which the alarm is generated as user **root**.

Step 4 View the $BIGDATA_LOG_HOME/nodeagent/scriptlog/checkfileconfig.log file and analyze the cause based on the error log. Locate the check standards of the file in the **Related Information** and manually check and modify the file based on the standards.

Run the **vi** *file name* command to enter the editing mode, and then press **Insert** to start editing.

After the modification is complete, press **Esc** to exit the editing mode and enter **:wq** to save the settings and exit.

For example:

vi /etc/ssh/sshd_config

Step 5 Wait a hour and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 7 Select **NodeAgent** from the **Service** and click **OK**.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

- **Check standards of /etc/fstab**

  Check whether the partitions configured in the **/etc/fstab** file can be found in **/proc/mounts**.

  Check whether the swap partitions configured in fstab correspond to those in /proc/swaps.

- **Check the /etc/hosts configuration file.**

  Run **cat /ect/hosts**. If any of the following situations occurs, the **/etc/hosts** configuration file is abnormal:

  a. The **/etc/hosts** file does not exist.
  b. The host name is not configured in the file.
  c. The host name maps to multiple IP addresses in the file.
  d. The IP address corresponding to the host name does not exist in the command output of the **ifconfig** command.
  e. One IP address maps to multiple host names in the file.

- **Check standards of /etc/ssh/sshd_config**

  Run the **vi /etc/ssh/sshd_config** command to check whether configuration items are configured as follows:

  a. The value of **UseDNS** must be set to **no**.
  b. The value of **MaxStartups** must be greater than or equal to 1000.
  c. At least one of the **PasswordAuthentication** and **ChallengeResponseAuthentication** parameters must be left blank or at least one of the parameters be set to **yes**.

9.13.25 ALM-12045 Read Packet Dropped Rate Exceeds the Threshold

Description

The system checks the read packet dropped rate every 30 seconds. This alarm is generated when the read packet dropped rate exceeds the threshold (the default threshold is 0.5%) for multiple times (the default value is **5**).

To change the threshold, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Reading** > **Read Packet Dropped Rate**.

This alarm is cleared when **Trigger Count** is 1 and the read packet dropped rate is less than or equal to the threshold. This alarm is cleared when **Trigger Count** is greater than 1 and the read packet dropped rate is less than or equal to 90% of the threshold.

The alarm detection is disabled by default. If you want to enable this function, check whether this function can be enabled based on Checking System Environments.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12045	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
PortName	Specifies the network port for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

The service performance deteriorates or some services time out.

Risk warning: In SUSE kernel 3.0 or later or Red Hat 7.2, the system kernel modifies the mechanism for counting the number of dropped read packets. In this case, this alarm may be generated even if the network is running properly, but services are not affected. You are advised to check the system environment first.

Possible Causes

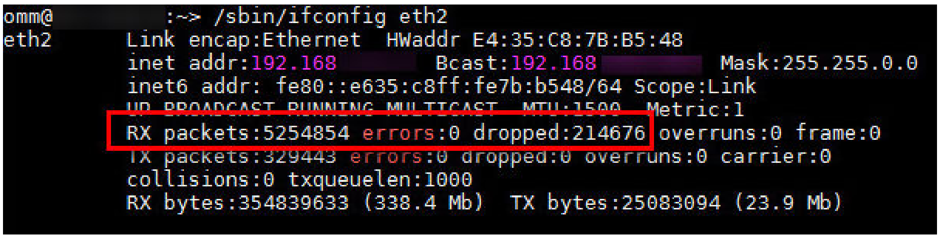
- An OS exception occurs.
- The NICs are bonded in active/standby mode.
- The alarm threshold is improperly configured.
- The network quality is poor.

Procedure

View the network packet dropped rate.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. On the page that is displayed, click ⌄ in the row containing the alarm, and view the name of the host for which the alarm is generated and the NIC name.

Step 2 Log in to the alarm node as user **omm**, and run the **/sbin/ifconfig** *NIC name* command to check whether packet loss occurs on the network.



```
omm@[redacted]:~> /sbin/ifconfig eth2
eth2      Link encap:Ethernet  HWaddr E4:35:C8:7B:B5:48
          inet addr:192.168[redacted]  Bcast:192.168[redacted]  Mask:255.255.0.0
          inet6 addr: fe80::e635:c8ff:fe7b:b548/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:5254854 errors:0 dropped:214676 overruns:0 frame:0
          TX packets:329443 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:354839633 (338.4 Mb)  TX bytes:25083094 (23.9 Mb)
```

NOTE

- *IP address of the node for which the alarm is generated*: Query the IP address of the node for which the alarm is generated on the **Hosts** page of FusionInsight Manager based on the value of **HostName** in the alarm location information. Check both the IP addresses of the management plane and service plane.
- Packet loss rate = (Number of dropped packets/Total number of received packets) x 100%. If the packet loss rate is greater than the system threshold (0.5% by default), read packets are dropped.

- If yes, go to **Step 11**.
- If no, go to **Step 3**.

Check the system environment.

Step 3 Log in to the active OMS node or the alarm node as user **omm**.

Step 4 Run the **cat /etc/*-release** command to check the OS type.

- For Red Hat Enterprise Linux, go to **Step 5**.

```
# cat /etc/*-release
Red Hat Enterprise Linux Server release 7.2 (Santiago)
```

- For SUSE Linux, go to **Step 6**.

```
# cat /etc/*-release
SUSE Linux Enterprise Server 11 (x86_64)
VERSION = 11
PATCHLEVEL = 3
```

- For other OS types, go to **Step 11**.

Step 5 Run the **cat /etc/redhat-release** command to check whether the OS version is **Red Hat 7.2 (x86)** or **Red Hat 7.4 (TaiShan)**.

```
# cat /etc/redhat-release
Red Hat Enterprise Linux Server release 7.2 (Santiago)
```

- If yes, the alarm sending function cannot be enabled. Go to **Step 7**.
- If no, go to **Step 11**.

Step 6 Run the **cat /proc/version** command to check whether the SUSE kernel version is 3.0 or later.

```
# cat /proc/version
Linux version 3.0.101-63-default (geeko@buildhost) (gcc version 4.3.4 [gcc-4_3-branch revision 152973] (SUSE Linux) ) #1 SMP Tue Jun 23 16:02:31 UTC 2015 (4b89d0c)
```

- If yes, the alarm sending function cannot be enabled. Go to **Step 7**.

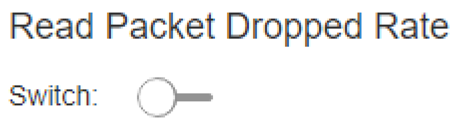
- If no, go to **Step 11**.

Step 7 Log in to FusionInsight Manager and choose **O&M** > **Alarm** > **Threshold Configuration**.

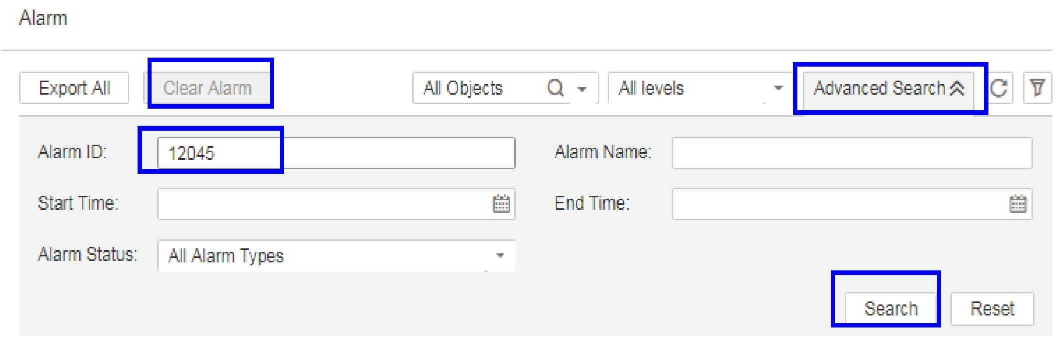
Step 8 In the navigation tree of the **Thresholds** page, choose *Name of the desired cluster* > **Host** > **Network Reading** > **Read Packet Dropped Rate**. In the area on the right, check whether the **Switch** is toggled on.

- If yes, the alarm sending function is enabled. Go to **Step 9**.
- If no, the alarm sending function is disabled. Go to **Step 10**.

Step 9 In the area on the right, toggle **Switch** off to disable the checking of **Network Read Packet Dropped Rate Exceeds the Threshold**.



Step 10 On the **Alarm** page of FusionInsight Manager, search for alarm **12045** and manually clear the alarm if it is not automatically cleared. No further action is required.



NOTE

ID of the Network Read Packet Dropped Rate Exceeds the Threshold alarm is **12045**.

Check whether the NICs are bonded in active/standby mode.

Step 11 Log in to the alarm node as user **omm** and run the **ls -l /proc/net/bonding** command to check whether the **/proc/net/bonding** directory exists on the node.

- If yes, the bond mode is configured for the node. Go to **Step 12**.
```
# ls -l /proc/net/bonding/
total 0
-r--r--r-- 1 root root 0 Oct 11 17:35 bond0
```
- If no, the bond mode is not configured for the node. Go to **Step 14**.
```
# ls -l /proc/net/bonding/
ls: cannot access /proc/net/bonding/: No such file or directory
```

Step 12 Run the **cat /proc/net/bonding/***bond0* command to check whether the value of **Bonding Mode** in the configuration file is **fault-tolerance**.

NOTE

In the command, **bond0** indicates the name of the bond configuration file. Use the file name obtained in **Step 11**.

```
# cat /proc/net/bonding/bond0
Ethernet Channel Bonding Driver: v3.7.1 (April 27, 2011)

Bonding Mode: fault-tolerance (active-backup)
Primary Slave: eth1 (primary_reselect always)
Currently Active Slave: eth1
MII Status: up
MII Polling Interval (ms): 100
Up Delay (ms): 0
Down Delay (ms): 0

Slave Interface: eth0
MII Status: up
Speed: 1000 Mbps
Duplex: full
Link Failure Count: 1
Slave queue ID: 0

Slave Interface: eth1
MII Status: up
Speed: 1000 Mbps
Duplex: full
Link Failure Count: 1
Slave queue ID: 0
```

- If yes, the NICs are bonded in active/standby mode. Go to **Step 13**.
- If no, go to **Step 14**.

Step 13 Check whether the NIC specified by **NetworkCardName** in the alarm is the standby NIC.

- If yes, the alarm of the standby NIC cannot be automatically cleared. Manually clear the alarm on the alarm management page. No further action is required.
- If no, go to **Step 14**.

> NOTE
>
> To determine the standby NIC, check the **/proc/net/bonding/bond0** configuration file. If the NIC name corresponding to **NetworkCardName** is **Slave Interface** but not **Currently Active Slave** (the current active NIC), the NIC is the standby one.

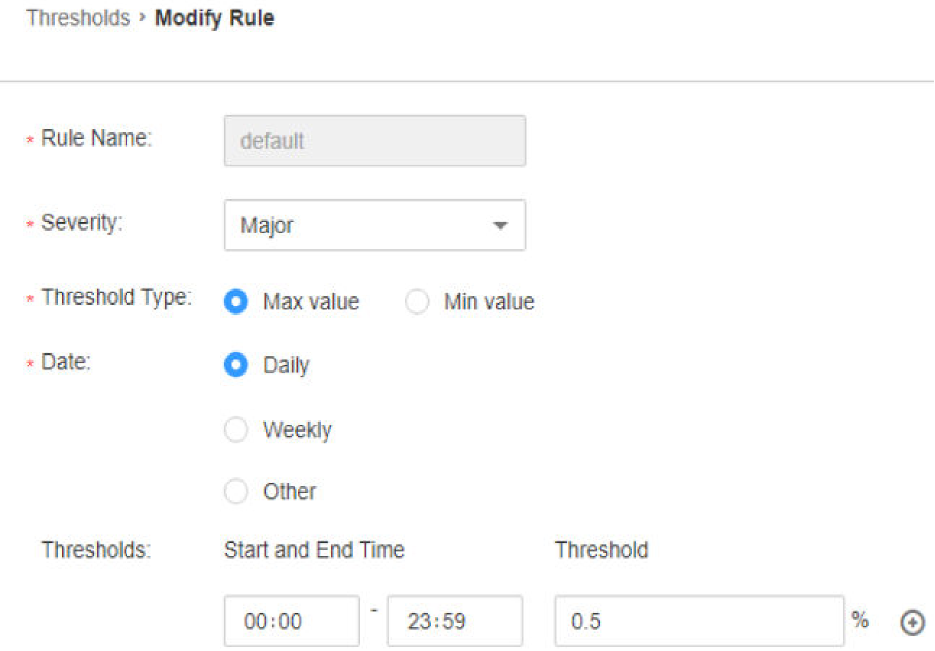
Check whether the threshold is set properly.

Step 14 Log in to FusionInsight Manager, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Reading** > **Read Packet Dropped Rate**, and check whether the alarm threshold is configured properly. The default value is **0.5%**. You can adjust the threshold as needed.

- If yes, go to **Step 17**.
- If no, go to **Step 15**.

Step 15 Choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Reading** > **Read Packet Dropped Rate**. Click **Modify** in the **Operation** column to change the threshold. See **Figure 9-18**.

Figure 9-18 Configuring the alarm threshold



Thresholds > **Modify Rule**

* Rule Name: default

* Severity: Major

* Threshold Type: ● Max value ○ Min value

* Date: ● Daily ○ Weekly ○ Other

Thresholds: Start and End Time 00:00 - 23:59 Threshold 0.5 %

Step 16 After 5 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 17**.

Check whether the network connection is normal.

Step 17 Contact the network administrator to check whether the network is normal.

- If yes, rectify the fault and go to **Step 18**.
- If no, go to **Step 19**.

Step 18 After 5 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 19**.

Collect the fault information.

Step 19 On FusionInsight Manager of the active cluster, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 20 Select **OMS** for **Service** and click **OK**.

Step 21 Expand the **Hosts** dialog box and select the alarm node and the active OMS node.

Step 22 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 23 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.26 ALM-12046 Write Packet Dropped Rate Exceeds the Threshold

Description

The system checks the write packet dropped rate every 30 seconds. This alarm is generated when the write packet dropped rate exceeds the threshold (the default threshold is 0.5%) for multiple times (the default value is **5**).

To change the threshold, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Writing** > **Write Packet Dropped Rate**.

If **Trigger Count** is **1**, this alarm is cleared when the network write packet dropped rate is less than or equal to the threshold. If **Trigger Count** is greater than **1**, this alarm is cleared when the network write packet dropped rate is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12046	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Port Name	Specifies the network port for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

The service performance deteriorates or some services time out.

Possible Causes

- The alarm threshold is improperly configured.
- The network quality is poor.

Procedure

Check whether the threshold is set properly.

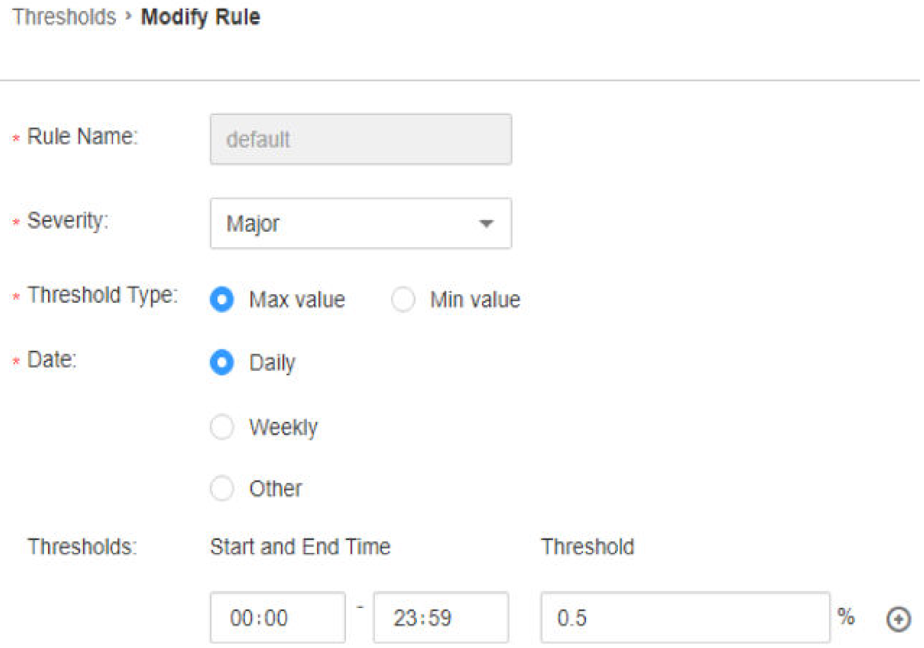
Step 1 Log in to FusionInsight Manager, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Writing** > **Write Packet Dropped Rate**, and check whether the alarm threshold is configured properly. The default value is **0.5%**. You can adjust the threshold as needed.

- If yes, go to **Step 4**.
- If no, go to **Step 2**.

Step 2 Choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Writing** > **Write Packet Dropped Rate**. Click **Modify** in the **Operation** column to change the threshold.

See **Figure 9-19**.

Figure 9-19 Configuring the alarm threshold



* Rule Name: default

* Severity: Major

* Threshold Type: ● Max value ○ Min value

* Date: ● Daily ○ Weekly ○ Other

Thresholds:

Start and End Time		Threshold
00:00	- 23:59	0.5 %

Step 3 After 5 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check whether the network connection is normal.

Step 4 Contact the network administrator to check whether the network is normal.

- If yes, rectify the fault and go to **Step 5**.
- If no, go to **Step 6**.

Step 5 After 5 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager of the active cluster, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Select **OMS** for **Service** and click **OK**.

Step 8 Expand the **Hosts** dialog box and select the alarm node and the active OMS node.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 10 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.27 ALM-12047 Read Packet Error Rate Exceeds the Threshold

Description

The system checks the read packet error rate every 30 seconds. This alarm is generated when the read packet error rate exceeds the threshold (the default threshold is **0.5%**) for multiple times (the default value is **5**).

To change the threshold, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Reading** > **Read Packet Error Rate**.

If **Trigger Count** is **1**, this alarm is cleared when the read packet error rate is less than or equal to the threshold. If **Trigger Count** is greater than **1**, this alarm is cleared when the read packet error rate is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12047	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Port Name	Specifies the network port for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

The communication is intermittently interrupted, and services time out.

Possible Causes

- The alarm threshold is improperly configured.
- The network quality is poor.

Procedure

Check whether the threshold is set properly.

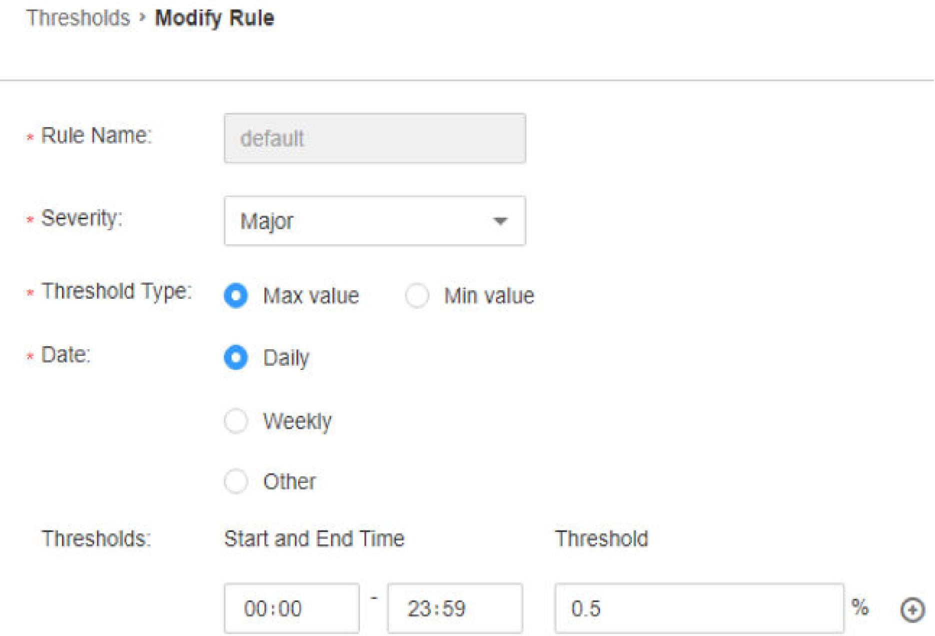
Step 1 Log in to FusionInsight Manager, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Reading** > **Read Packet Error Rate**, and check whether the alarm threshold is configured properly. The default value is **0.5%**. You can adjust the threshold as needed.

- If yes, go to **Step 4**.
- If no, go to **Step 2**.

Step 2 Choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Reading** > **Read Packet Error Rate**. Click **Modify** in the **Operation** column to change the threshold.

See **Figure 9-20**.

Figure 9-20 Configuring the alarm threshold



Thresholds › **Modify Rule**

* Rule Name: default

* Severity: Major

* Threshold Type: ● Max value ○ Min value

* Date: ● Daily
○ Weekly
○ Other

Thresholds:	Start and End Time	Threshold
	00:00 - 23:59	0.5 %

Step 3 After 5 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check whether the network connection is normal.

Step 4 Contact the network administrator to check whether the network is normal.

- If yes, rectify the fault and go to **Step 5**.
- If no, go to **Step 6**.

Step 5 After 5 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager of the active cluster, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Select **OMS** for **Service** and click **OK**.

Step 8 Expand the **Hosts** dialog box and select the alarm node and the active OMS node.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 10 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.28 ALM-12048 Write Packet Error Rate Exceeds the Threshold

Description

The system checks the write packet error rate every 30 seconds. This alarm is generated when the write packet error rate exceeds the threshold (the default threshold is **0.5%**) for multiple times (the default value is **5**).

To change the threshold, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Writing** > **Write Packet Error Rate**.

If **Trigger Count** is **1**, this alarm is cleared when the write packet error rate is less than or equal to the threshold. If **Trigger Count** is greater than **1**, this alarm is cleared when the write packet error rate is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12048	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Port Name	Specifies the network port for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

The communication is intermittently interrupted, and services time out.

Possible Causes

- The alarm threshold is improperly configured.
- The network quality is poor.

Procedure

Check whether the threshold is set properly.

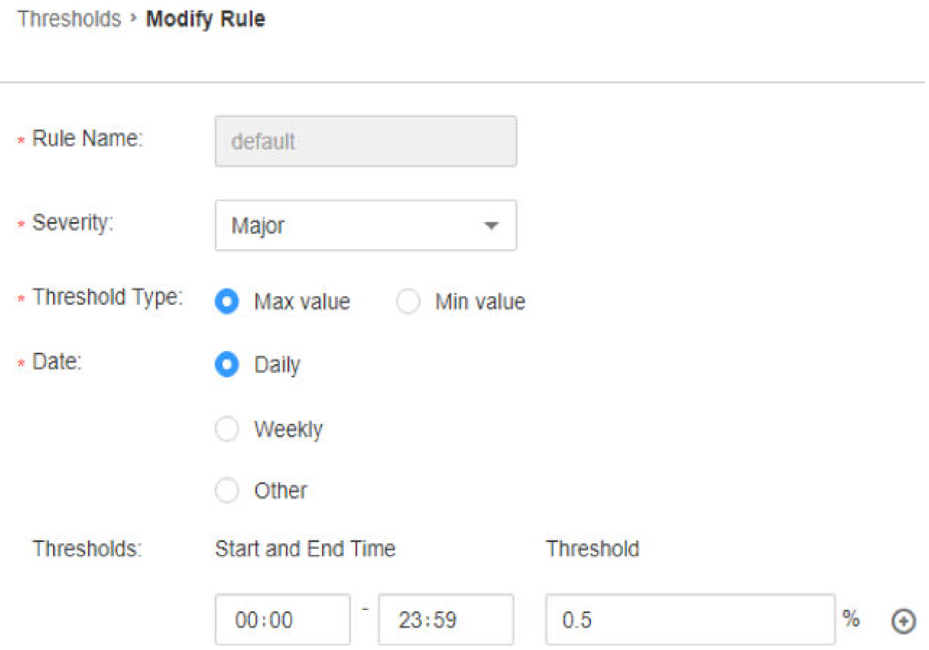
Step 1 Log in to FusionInsight Manager, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Writing** > **Write Packet Error Rate**, and check whether the alarm threshold is configured properly. The default value is **0.5%**. You can adjust the threshold as needed.

- If yes, go to **Step 4**.
- If no, go to **Step 2**.

Step 2 Choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Writing** > **Write Packet Error Rate**. Click **Modify** in the **Operation** column to change the threshold.

See **Figure 9-21**.

Figure 9-21 Configuring the alarm threshold



* Rule Name: default

* Severity: Major

* Threshold Type: ● Max value ○ Min value

* Date: ● Daily ○ Weekly ○ Other

Thresholds: Start and End Time 00:00 - 23:59 Threshold 0.5 %

Step 3 After 5 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check whether the network connection is normal.

Step 4 Contact the network administrator to check whether the network is normal.

- If yes, rectify the fault and go to **Step 5**.
- If no, go to **Step 6**.

Step 5 After 5 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager of the active cluster, choose **O&M** > **Log** > **Download**.

Step 7 Select **OMS** for **Service** and click **OK**.

Step 8 Expand the **Hosts** dialog box and select the alarm node and the active OMS node.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 10 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.29 ALM-12049 Network Read Throughput Rate Exceeds the Threshold

Description

The system checks the network read throughput rate every 30 seconds and compares the actual throughput rate with the threshold (the default threshold is 80%). This alarm is generated when the system detects that the network read throughput rate exceeds the threshold for several times (5 times by default) consecutively.

To change the threshold, choose **O&M > Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Reading** > **Read Throughput Rate**.

When the **Trigger Count** is 1, this alarm is cleared when the network read throughput rate is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the network read throughput rate is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12049	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
NetworkCardName	Specifies the network port for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

The service system runs improperly or is unavailable.

Possible Causes

- The alarm threshold is set improperly.
- The network port rate cannot meet the current service requirements.

Procedure

Check whether the threshold is set properly.

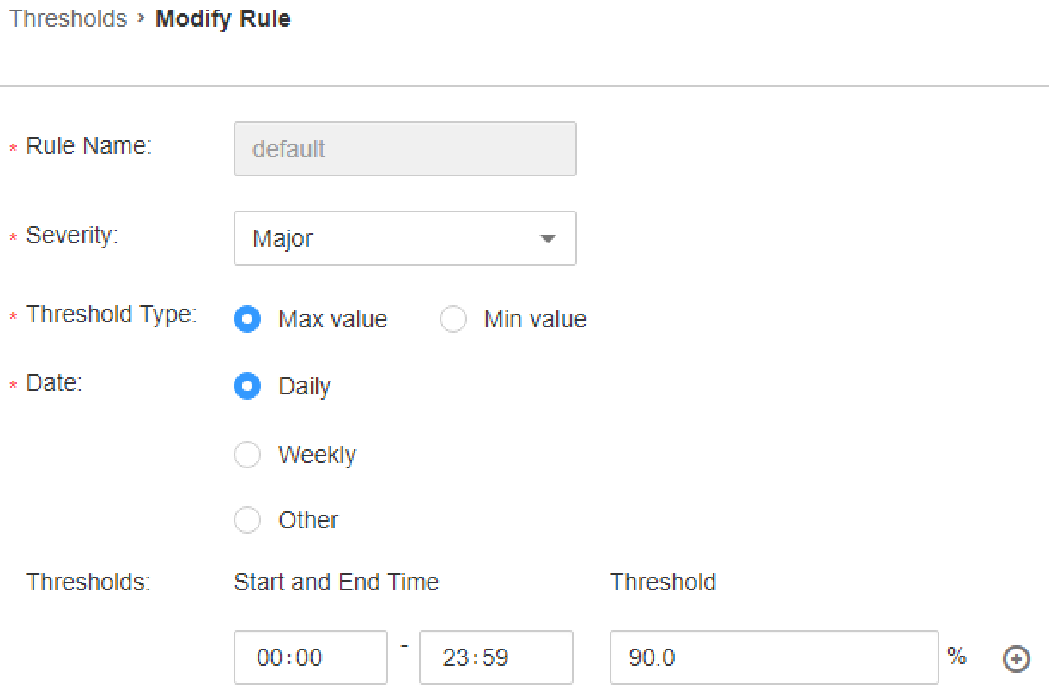
Step 1 On the FusionInsight Manager, choose **O&M > Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Reading** > **Read Throughput Rate** and check whether the alarm threshold is set properly. (By default, 80% is a proper value. However, users can configure the value as required.)

- If yes, go to **Step 2**.
- If no, go to **Step 4**.

Step 2 Based on actual usage condition, choose **O&M > Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Reading** > **Read Throughput Rate** and click **Modify** in the **Operation** column to modify the alarm threshold.

For details, see **Figure 9-22**.

Figure 9-22 Setting alarm thresholds



Thresholds › **Modify Rule**

* Rule Name: default

* Severity: Major

* Threshold Type: ● Max value ○ Min value

* Date: ● Daily ○ Weekly ○ Other

Thresholds: Start and End Time: 00:00 - 23:59 Threshold: 90.0 %

Step 3 Wait for 5 minutes, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check whether the network port rate can meet the service requirements.

Step 4 On FusionInsight Manager, click ∨ in the row where the alarm is located in the real-time alarm list and obtain the IP address of the host and the network port name for which the alarm is generated.

Step 5 Log in to the host for which the alarm is generated as user **root**.

Step 6 Run the **ethtool** *network port name* command to check the maximum speed of the current network port.

NOTE

In the VM environment, you cannot run a command to query the network port rate. It is recommended that you contact the system administrator to confirm whether the network port rate meets the requirements.

Step 7 If the network read throughput rate exceeds the threshold, contact the system administrator to increase the network port rate.

Step 8 Check whether the alarm is cleared.

- If yes, no further action is required.

- If no, go to **Step 9**.

Collect fault information.

Step 9 On the FusionInsight Manager home page of the active cluster, choose **O&M** > **Log** > **Download**.

Step 10 Select **OMS** from the **Service** and click **OK**.

Step 11 Set **Host** to the node for which the alarm is generated and the active OMS node.

Step 12 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 13 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.30 ALM-12050 Network Write Throughput Rate Exceeds the Threshold

Description

The system checks the network write throughput rate every 30 seconds and compares the actual throughput rate with the threshold (the default threshold is 80%). This alarm is generated when the system detects that the network write throughput rate exceeds the threshold for several times (5 times by default) consecutively.

To change the threshold, choose **O&M > Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Writing** > **Write Throughput Rate**.

When the **Trigger Count** is 1, this alarm is cleared when the network write throughput rate is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the network write throughput rate is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12050	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
NetworkCardName	Specifies the network port for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

The service system runs improperly or is unavailable.

Possible Causes

- The alarm threshold is set improperly.
- The network port rate cannot meet the current service requirements.

Procedure

Check whether the threshold is set properly.

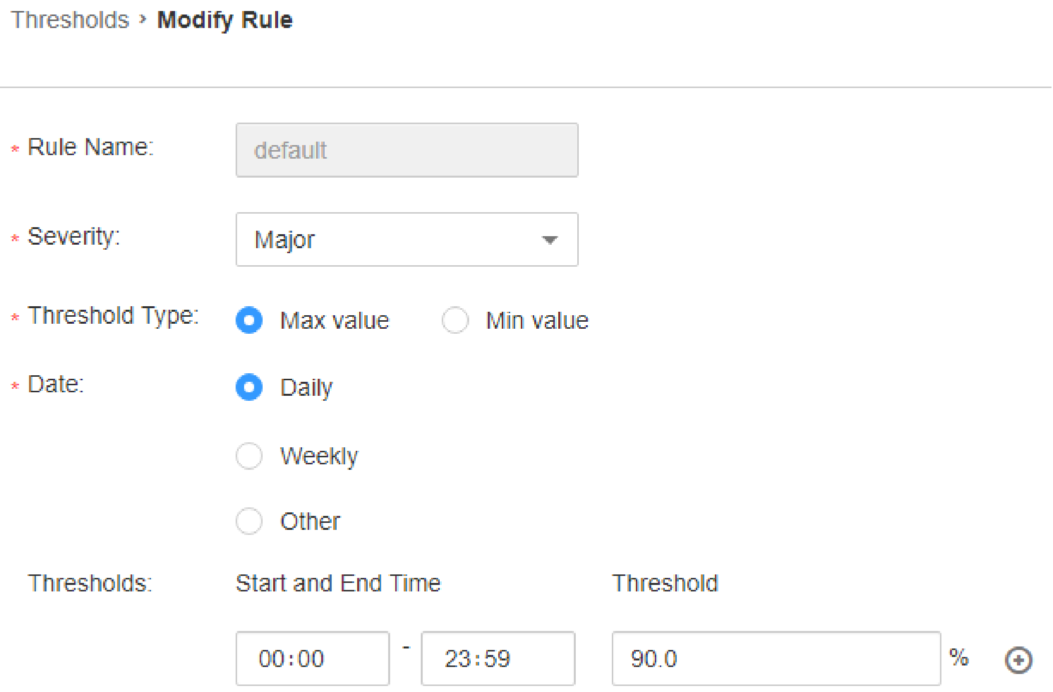
Step 1 On the FusionInsight Manager, choose **O&M > Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Writing** > **Write Throughput Rate** and check whether the alarm threshold is set properly. (By default, 80% is a proper value. However, users can configure the value as required.)

- If yes, go to **Step 4**.
- If no, go to **Step 2**.

Step 2 Based on actual usage condition, choose **O&M > Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Writing** > **Write Throughput Rate** and click **Modify** in the **Operation** column to modify the alarm threshold.

For details, see **Figure 9-23**.

Figure 9-23 Setting alarm thresholds



Thresholds › **Modify Rule**

* Rule Name:	default
* Severity:	Major
* Threshold Type:	● Max value ○ Min value
* Date:	● Daily ○ Weekly ○ Other

Thresholds: Start and End Time: 00:00 - 23:59 Threshold: 90.0 %

Step 3 Wait for 5 minutes, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check whether the network port rate can meet the service requirements.

Step 4 On FusionInsight Manager, click ∨ in the row where the alarm is located in the real-time alarm list and obtain the IP address of the host and the network port name for which the alarm is generated.

Step 5 Log in to the host for which the alarm is generated as user **root**.

Step 6 Run the **ethtool***network port name* command to check the maximum speed of the current network port.

NOTE

In the VM environment, you cannot run a command to query the network port rate. It is recommended that you contact the system administrator to confirm whether the network port rate meets the requirements.

Step 7 If the network write throughput rate exceeds the threshold, contact the system administrator to increase the network port rate.

Step 8 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect fault information.

Step 9 On the FusionInsight Manager home page of the active cluster, choose **O&M** > **Log** > **Download**.

Step 10 Select **OMS** from the **Service** and click **OK**.

Step 11 Set **Host** to the node for which the alarm is generated and the active OMS node.

Step 12 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 13 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.31 ALM-12051 Disk Inode Usage Exceeds the Threshold

Description

The system checks the disk Inode usage every 30 seconds and compares the actual Inode usage with the threshold (the default threshold is 80%). This alarm is generated when the Inode usage exceeds the threshold for several times (5 times by default) consecutively.

To change the threshold, choose **O&M > Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Disk** > **Disk Inode Usage**.

When the **Trigger Count** is 1, this alarm is cleared when the disk Inode usage is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the disk Inode usage is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12051	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.

Name	Meaning
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
PartitionName	Specifies the disk partition for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

Data cannot be properly written to the file system.

Possible Causes

Massive small files are stored in the disk.

Procedure

Massive small files are stored in the disk.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** and click ⌄ in the row where the alarm is located in the real-time alarm list and obtain the IP address of the host and the disk partition for which the alarm is generated.

Step 2 Log in to the host for which the alarm is generated as user **root**.

Step 3 Run the **df -i | grep -iE "** *partition name*|Filesystem" command to check the current disk Inode usage.

```
# df -i | grep -iE "xvda2|Filesystem"
Filesystem           Inodes   IUsed   IFree IUse% Mounted on
/dev/xvda2           2359296  207420 2151876    9% /
```

Step 4 If the Inode usage exceeds the threshold, manually check small files stored in the disk partition and confirm whether these small files can be deleted.

NOTE

Run the **for i in /*; do echo $i; find $i|wc -l; done** command to query the number of files in a partition. Replace **/*** with the specified partition.

```
# for i in /srv/*; do echo $i; find $i|wc -l; done
/srv/BigData
4284
/srv/ftp
1
/srv/www
13
```

- If yes, run the **rm -rf** *Path of the file or folder* to be deleted command to delete the file or folder and go to **Step 5**.

NOTE

Deleting a file or folder is a high-risk operation. Ensure that the file or folder is no longer required before performing this operation.

- If no, expand the capacity. Then, perform **Step 5**.

Step 5 Wait for 5 minutes, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager home page of the active cluster, choose **O&M** > **Log** > **Download**.

Step 7 Select **OMS** from the **Service** and click **OK**.

Step 8 Set **Host** to the node for which the alarm is generated and the active OMS node.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.32 ALM-12052 TCP Temporary Port Usage Exceeds the Threshold

Description

The system checks the TCP temporary port usage every 30 seconds and compares the actual usage with the threshold (the default threshold is 80%). This alarm is generated when the TCP temporary port usage exceeds the threshold for several times (5 times by default) consecutively.

To change the threshold, choose **O&M > Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Network Status** > **TCP Ephemeral Port Usage**.

When the **Trigger Count** is 1, this alarm is cleared when the TCP temporary port usage is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the TCP temporary port usage is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12052	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

Services on the host cannot establish external connections, and therefore they are interrupted.

Possible Causes

- The temporary port cannot meet the current service requirements.
- The system is abnormal.

Procedure

Expand the temporary port number range.

Step 1 On FusionInsight Manager, click ⌄ in the row where the alarm is located in the real-time alarm list and obtain the IP address of the host for which the alarm is generated.

Step 2 Log in to the host for which the alarm is generated as user **omm**.

Step 3 Run the **cat /proc/sys/net/ipv4/ip_local_port_range |cut -f 1** command to obtain the value of the start port and run the **cat /proc/sys/net/ipv4/ip_local_port_range |cut -f 2** command to obtain the value of the end port. The total number of temporary ports is the value of the end port minus the value of the start port. If the total number of temporary ports is smaller than 28,232, the random port range of the OS is narrow. Contact the system administrator to increase the port range.

Step 4 Run the **ss -ant 2>/dev/null | grep -v LISTEN | awk 'NR > 2 {print $4}'|cut -d ':' -f 2 | awk '$1 >"***Value of the start port***" {print $1}' | sort -u | wc -l** command to calculate the number of used temporary ports.

Step 5 The formula for calculating the usage of the temporary ports is: Usage of the temporary ports = (Number of used temporary ports/Total number of temporary ports) x 100%. Check whether the temporary port usage exceeds the threshold.

- If yes, go to **Step 7**.
- If no, go to **Step 6**.

Step 6 Wait for 5 minutes, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Check whether the system environment is abnormal.

Step 7 Run the following command to import the temporary file and view the frequently used ports in the **port_result.txt file**:

netstat -tnp|sort > $BIGDATA_HOME/tmp/port_result.txt

```
netstat -tnp|sort

Active Internet connections (w/o servers)

Proto Recv Send LocalAddress ForeignAddress State PID/ProgramName tcp   0   0 10-120-85-154:45433
10-120-85-154:9866 CLOSE_WAIT 94237/java
tcp   0   0 10-120-85-154:45434  10-120-85-154:9866 CLOSE_WAIT 94237/java
tcp   0   0 10-120-85-154:45435  10-120-85-154:9866 CLOSE_WAIT 94237/java
...
```

Step 8 Run the following command to view the processes that occupy a large number of ports:

ps -ef |grep *PID*

NOTE

- PID is the processes ID queried in **Step 7**.
- Run the following command to collect information about all processes and check the processes that occupy a large number of ports:

  ps -ef > $BIGDATA_HOME/tmp/ps_result.txt

Step 9 After obtaining the administrator's approval, clear the processes that occupy a large number of ports. Wait for 5 minutes, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 10**.

Collect fault information.

Step 10 On the FusionInsight Manager home page of the active cluster, choose **O&M** > **Log** > **Download**.

Step 11 Select **OMS** from the **Service** and click **OK**.

Step 12 Set **Host** to the node for which the alarm is generated and the active OMS node.

Step 13 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 14 Contact the O&M personnel and send the collected log information and files **port_result.txt** and **ps_result.txt**. Then, delete the two residual temporary files from the environment.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.33 ALM-12053 Host File Handle Usage Exceeds the Threshold

Description

The system checks the file handle usage every 30 seconds and compares the actual usage with the threshold (the default threshold is 80%). This alarm is generated when the host file handle usage exceeds the threshold for several times (5 times by default) consecutively.

To change the threshold, choose **O&M > Alarm** > **Thresholds** > *Name of the desired cluster* > **Host** > **Host Status** > **Host File Handle Usage**.

When the **Trigger Count** is 1, this alarm is cleared when the host file handle usage is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the host file handle usage is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12053	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

The I/O operations, such as opening a file or connecting to network, cannot be performed and programs are abnormal.

Possible Causes

- The application process is abnormal. For example, the opened file or socket is not closed.
- The number of file handles cannot meet the current service requirements.
- The system is abnormal.

Procedure

Check information about files opened in processes.

Step 1 On FusionInsight Manager, click ⌄ in the row where the alarm is located in the real-time alarm list and obtain the IP address of the host for which the alarm is generated.

Step 2 Log in to the host for which the alarm is generated as user **root**.

Step 3 Run the **lsof -n|awk '{print $2}'|sort|uniq -c|sort -nr|more** command to check the process that occupies excessive file handles.

Step 4 Check whether the processes in which a large number of files are opened are normal. For example, check whether there are files or sockets not closed.

- If yes, go to **Step 5**.
- If no, go to **Step 7**.

Step 5 Release the abnormal processes that occupy too many file handles.

Step 6 Five minutes later, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Increase the number of file handles.

Step 7 On FusionInsight Manager, click ⌄ in the row where the alarm is located in the real-time alarm list and obtain the IP address of the host for which the alarm is generated.

Step 8 Log in to the host for which the alarm is generated as user **root**.

Step 9 Contact the system administrator to increase the number of system file handles.

Step 10 Run the **cat /proc/sys/fs/file-nr** command to view the used handles and the maximum number of file handles. The first value is the number of used handles, the third value is the maximum number. Please check whether the usage exceeds the threshold.

- If yes, go to **Step 9**.
- If no, go to **Step 11**.

```
# cat /proc/sys/fs/file-nr
12704 0 640000
```

Step 11 Wait for 5 minutes, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 12**.

Check whether the system environment is abnormal.

Step 12 Contact the system administrator to check whether the operating system is abnormal.

- If yes, go to **Step 13** to rectify the fault.
- If no, go to **Step 14**.

Step 13 Wait for 5 minutes, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 14**.

Collect fault information.

Step 14 On the FusionInsight Manager home page of the active cluster, choose **O&M** > **Log** > **Download**.

Step 15 Select **OMS** from the **Service** and click **OK**.

Step 16 Set **Host** to the node for which the alarm is generated and the active OMS node.

Step 17 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 18 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.34 ALM-12054 Invalid Certificate File

Description

The system checks whether the certificate file is invalid (has expired or is not valid yet) on 23:00 every day. This alarm is generated when the certificate file is invalid.

This alarm is cleared when a valid certificate is imported.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12054	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Some functions are unavailable.

Possible Causes

No certificate (CA certificate, HA root certificate, HA user certificate, Gaussdb root certificate, or Gaussdb user certificate) is imported to the system, the certificate fails to be imported, or the certificate file is invalid.

Procedure

Check the alarm cause.

Step 1 On FusionInsight Manager, locate the target alarm in the real-time alarm list and click ⌄.

View **Additional Information** to obtain the additional information about the alarm.

- If **CA Certificate** is displayed in the additional alarm information, log in to the active OMS management node as user **omm** and go to **Step 2**.
- If **HA root Certificate** is displayed in the additional information, view **Location** to obtain the name of the host involved in this alarm. Then, log in to the host as user **omm** and go to **Step 3**.
- If **HA server Certificate** is displayed in the additional information, view **Location** to obtain the name of the host involved in this alarm. Then, log in to the host as user **omm** and go to **Step 4**.
- If **Certificate has expired** is displayed in the additional information, view **Location** to obtain the name of the host for which the alarm is generated. Then, log in to the host as user **omm** and perform **Step 2** to **Step 4** in sequence to check whether the certificates have expired. If these certificates have not expired, check whether other certificates have been imported. If yes, import the certificate files again.

Check the validity period of the certificate files in the system.

Step 2 Check whether the current system time is in the validity period of the CA certificate.

Run the **bash ${CONTROLLER_HOME}/security/cert/conf/querycertvalidity.sh** command to check the effective time and due time of the CA root certificate.

- If yes, go to **Step 7**.
- If no, go to **Step 5**.

Step 3 Check whether the current system time is in the validity period of the HA root certificate.

Run the **openssl x509 -noout -text -in ${CONTROLLER_HOME}/security/certHA/root-ca.crt** command to check the effective time and due time of the HA root certificate.

- If yes, go to **Step 7**.
- If no, go to **Step 6**.

Step 4 Check whether the current system time is in the validity period of the HA user certificate.

Run the **openssl x509 -noout -text -in ${CONTROLLER_HOME}/security/certHA/server.crt** command to check the effective time and due time of the HA user certificate.

- If yes, go to **Step 7**.
- If no, go to **Step 6**.

The following is an example of the effective time and due time of a CA or HA certificate:

```
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number:
            97:d5:0e:84:af:ec:34:d8
        Signature Algorithm: sha256WithRSAEncryption
        Issuer: C=CN, ST=xxx, L=yyy, O=zzz, OU=IT, CN=HADOOP.COM
        Validity
            Not Before: Dec 13 06:38:26 2016 GMT             // Effective time
            Not After : Dec 11 06:38:26 2026 GMT             // Due time
```

Import certificate files.

Step 5 Import a new CA certificate file.

Apply for or generate a new CA certificate file and import it to the system. The alarm is automatically cleared after the CA certificate is imported. Check whether this alarm is reported again during periodic check.

- If yes, go to **Step 7**.
- If no, no further action is required.

Step 6 Import a new HA certificate file.

Apply for or generate a new HA certificate file and import it to the system. The alarm is automatically cleared after the CA certificate is imported. Check whether this alarm is reported again during periodic check.

- If yes, go to **Step 7**.
- If no, no further action is required.

Collect the fault information.

Step 7 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 8 In the **Services** area, select **Controller**, **OmmServer**, **OmmCore**, and **Tomcat**, and click **OK**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.35 ALM-12055 The Certificate File Is About to Expire

Description

The system checks the certificate file on 23:00 every day. This alarm is generated if the certificate file is about to expire within 30 days.

This alarm is cleared when a certificate that is not about to expire is imported.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12055	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Some functions are unavailable.

Possible Causes

The remaining validity period of a system certificate (CA certificate, HA root certificate, HA user certificate, Gaussdb root certificate, or Gaussdb user certificate) is less than 30 days.

Procedure

Check the alarm cause.

Step 1 On FusionInsight Manager, locate the target alarm in the real-time alarm list and click ⌄.

View **Additional Information** to obtain the additional information about the alarm.

- If **CA Certificate** is displayed in the additional alarm information, log in to the active OMS management node as user **omm** and go to **Step 2**.
- If **HA root Certificate** is displayed in the additional information, view **Location** to obtain the name of the host involved in this alarm. Then, log in to the host as user **omm** and go to **Step 3**.

- If **HA server Certificate** is displayed in the additional information, view **Location** to obtain the name of the host involved in this alarm. Then, log in to the host as user **omm** and go to **Step 4**.

Check the validity period of the certificate files in the system.

Step 2 Check whether the remaining validity period of the CA certificate is smaller than the alarm threshold.

Run the **bash ${CONTROLLER_HOME}/security/cert/conf/querycertvalidity.sh** command to check the effective time and due time of the CA root certificate.

- If yes, go to **Step 5**.
- If no, go to **Step 7**.

Step 3 Check whether the remaining validity period of the HA root certificate is smaller than the alarm threshold.

Run the **openssl x509 -noout -text -in ${CONTROLLER_HOME}/security/certHA/root-ca.crt** command to check the effective time and due time of the HA root certificate.

- If yes, go to **Step 6**.
- If no, go to **Step 7**.

Step 4 Check whether the remaining validity period of the HA user certificate is smaller than the alarm threshold.

Run the **openssl x509 -noout -text -in ${CONTROLLER_HOME}/security/certHA/server.crt** command to check the effective time and due time of the HA user certificate.

- If yes, go to **Step 6**.
- If no, go to **Step 7**.

The following is an example of the effective time and due time of a CA or HA certificate:

```
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number:
            97:d5:0e:84:af:ec:34:d8
        Signature Algorithm: sha256WithRSAEncryption
        Issuer: C=CN, ST=xxx, L=yyy, O=zzz, OU=IT, CN=HADOOP.COM
        Validity
            Not Before: Dec 13 06:38:26 2016 GMT            // Effective time
            Not After : Dec 11 06:38:26 2026 GMT            // Due time
```

Import certificate files.

Step 5 Import a new CA certificate file.

Apply for or generate a new CA certificate file and import it to the system. Manually clear the alarm and check whether this alarm is generated again during periodic check.

- If yes, go to **Step 7**.
- If no, no further action is required.

Step 6 Import a new HA certificate file.

Apply for or generate a new HA certificate file and import it to the system. Manually clear the alarm and check whether this alarm is generated again during periodic check.

- If yes, go to **Step 7**.
- If no, no further action is required.

Collect the fault information.

Step 7 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 8 In the **Services** area, select **Controller**, **OmmServer**, **OmmCore**, and **Tomcat**, and click **OK**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.36 ALM-12057 Metadata Not Configured with the Task to Periodically Back Up Data to a Third-Party Server

Description

After the system is installed, it checks whether the task for periodically backing up metadata to the third-party server, and then performs the check hourly. If the task for periodically backing up metadata to a third-party server is not configured, a critical alarm is generated.

This alarm is cleared when a user creates such a backup task.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12057	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

If metadata is not backed up to a third-party server, metadata cannot be restored if both the active and standby management nodes of the cluster are faulty and local backup data is lost.

Possible Causes

Metadata is not configured with the task to periodically back up data to a third-party server.

Procedure

Step 1 On the FusionInsight Manager portal choose **O&M > Alarm > Alarms**.

Step 2 In the alarm list, click ∨ in the row where the alarm is located and identify the data module from which the alarm is generated based on **Additional Information**.

Step 3 Choose **O&M** > **Backup and Restoration > Backup Management** > **Create**.

Step 4 Configure a backup task. The backup data to be configured is consistent with the data in Additional Information of the alarm.

Step 5 After the backup task is created successfully, wait for two minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information

Step 6 On FusionInsight Manager, choose **O&M** > **Log > Download**.

Step 7 In the **Service** area, select **Controller** and click **OK**.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.37 ALM-12061 Process Usage Exceeds the Threshold

Description

The system checks the usage of the omm process every 30 seconds. Users can run the **ps -o nlwp, pid, args, -u omm | awk '{sum+=$1} END {print "", sum}'** command to obtain the number of concurrent processes of user **omm**. Run the **ulimit -u**command to obtain the maximum number of processes that can be simultaneously opened by user **omm**. Divide the number of concurrent processes by the maximum number to obtain the process usage of user **omm**. The process usage has a default threshold. This alarm is generated when the process usage exceeds the threshold.

If **Trigger Count** is **3** and the process usage is less than or equal to the threshold, this alarm is cleared. If **Trigger Count** is greater than **1**and the process usage is less than or equal to 90% of the threshold, this alarm is cleared.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12061	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Name	Meaning
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

- Switch to user **omm** fails.
- New omm process cannot be created.

Possible Causes

- The alarm threshold is improperly configured.
- The maximum number of processes (including threads) that can be concurrently opened by user **omm** is inappropriate.
- An excessive number of threads are opened at the same time.

Procedure

Check whether the alarm threshold or alarm hit number is properly configured.

Step 1 On the FusionInsight Manager, change the alarm threshold and **Trigger Count** based on the actual CPU usage.

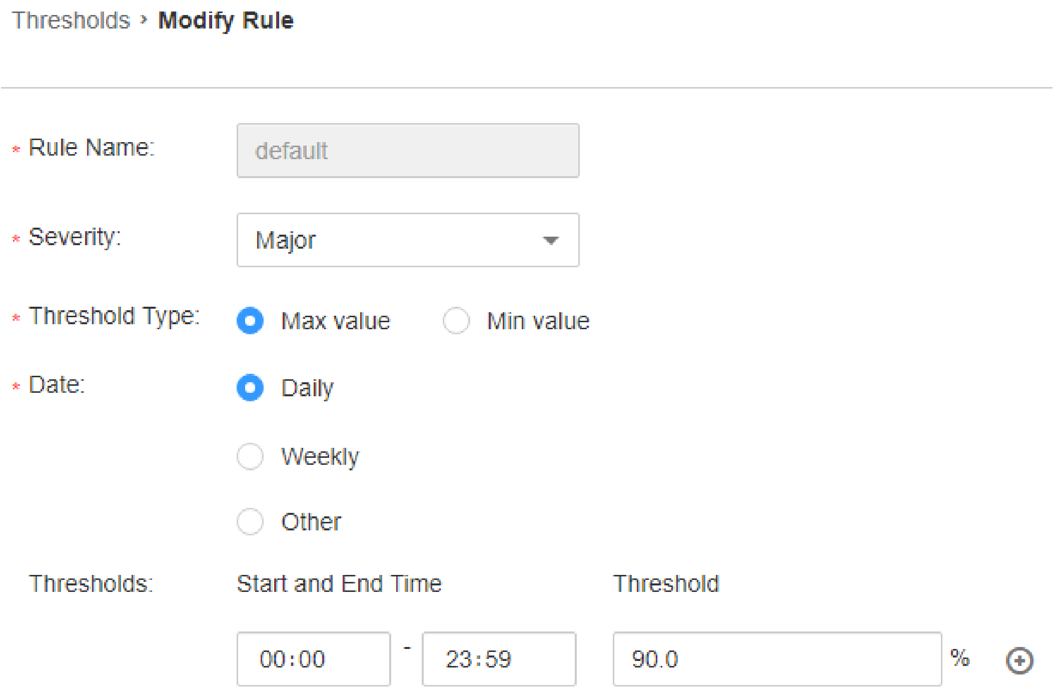
Specifically, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host**> **Process** > **omm Process Usage** to change Trigger Count.

NOTE

The alarm is generated when the process usage exceeds the threshold for the times specified by **Trigger Count**.

Set the alarm threshold based on the actual process usage. To check the process usage, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Host**> **Process** > **omm Process Usage**, as shown in **Figure 9-24**.

Figure 9-24 Setting an alarm threshold



Thresholds › **Modify Rule**

* Rule Name: default

* Severity: Major

* Threshold Type: ● Max value ○ Min value

* Date: ● Daily ○ Weekly ○ Other

Thresholds: Start and End Time 00:00 - 23:59 Threshold 90.0 %

Step 2 2 minutes later, check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 3**.

Check whether the maximum number of processes (including threads) opened by user omm is appropriate.

Step 3 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and view the IP address of the host for which the alarm is generated.

Step 4 Log in to the host where the alarm is generated as user **root**.

Step 5 Run the **su - omm** command to switch to user **omm**.

Step 6 Run the **ulimit -u** command to obtain the maximum number of threads that can be concurrently opened by user **omm** and check whether the number is greater than or equal to 60000.

- If it is, go to **Step 8**.
- If it is not, go to **Step 7**.

Step 7 Run the **ulimit -u 60000** command to change the maximum number to 60000. Two minutes later, check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 12**.

Check whether an excessive number of processes are opened at the same time.

Step 8 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and view the IP address of the host for which the alarm is generated.

Step 9 Log in to the host where the alarm is generated as user **root**.

Step 10 Run the **ps -o nlwp, pid, lwp, args, -u omm|sort -n** command to check the numbers of threads used by the system. The result is sorted based on the thread number. Analyze the top 5 thread numbers and check whether the threads are incorrectly used. If they are, contact maintenance personnel to rectify the fault. If they are not, run the **ulimit -u** command to change the maximum number to be greater than 60000.

Step 11 Five minutes later, check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 12**.

Collect fault information.

Step 12 On the FusionInsight Manager home page of the active clusters, choose **O&M** > **Log** > **Download**.

Step 13 Select **OmmServer** and **NodeAgent** from the **Service** and click **OK**.

Step 14 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 10 minutes before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 15 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.38 ALM-12062 OMS Parameter Configurations Mismatch with the Cluster Scale

Description

The system checks whether the OMS parameter configurations match with the cluster scale at each top hour. If the OMS parameter configurations do not meet the cluster scale requirements, the system generates this alarm. This alarm is automatically cleared when the OMS parameter configurations are modified.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12062	Major	Yes

Parameters

Parameter	Description
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the name of the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The OMS configuration is not modified when the cluster is installed or the system capacity is expanded.

Possible Causes

The OMS parameter configurations mismatch with the cluster scale.

Procedure

Check whether the OMS parameter configurations match with the cluster scale.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and view the IP address of the host for which the alarm is generated.

Step 2 Log in to the host where the alarm is generated as user **root**.

Step 3 Run the **su - omm** command to switch to user **omm**.

Step 4 Run the **vi $BIGDATA_LOG_HOME/controller/scriptlog/modify_manager_param.log** command to open the log file and search for the log file containing the following information: Current oms configurations cannot support *xx* nodes. In the information, *xx* indicates the number of nodes in the cluster.

Step 5 Optimize the current cluster configuration by following the instructions in **Optimizing Manager Configurations Based on the Number of Cluster Nodes**.

Step 6 One hour later, check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 7**.

Collect fault information.

Step 7 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 8 Select **Controller** from the **Service** and click **OK**.

Step 9 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

Optimizing Manager Configurations Based on the Number of Cluster Nodes

Step 1 Log in to the active Manager node as user **omm**.

Step 2 Run the following command to switch the directory:

cd ${BIGDATA_HOME}/om-server/om/sbin

Step 3 Run the following command to view the current Manager configurations.

sh oms_config_info.sh -q

Step 4 Run the following command to specify the number of nodes in the current cluster.

Command format: **sh oms_config_info.sh -s** *number of nodes*

Example:

sh oms_config_info.sh -s 1000

Enter **y** as prompted.

```
The following configurations will be modified:
    Module       Parameter         Current          Target
    Controller   controller.Xmx    4096m        =>  16384m
    Controller   controller.Xms    1024m        =>  8192m
    Controller   controller.node.heartbeat.error.threshold    30000                   =>   60000
    Pms          pms.mem           8192m        =>  10240m
Do you really want to do this operation? (y/n):
```

The configurations are updated successfully if the following information is displayed:

```
...
Operation has been completed. Now restarting OMS server.                 [done]
Restarted oms server successfully.
```

NOTE

- OMS is automatically restarted during the configuration update process.
- Clusters with similar quantities of nodes have same Manager configurations. For example, when the number of nodes is changed from 100 to 101, no configuration item needs to be updated.

----End

9.13.39 ALM-12063 Unavailable Disk

Description

The system checks whether the data disk of the current host is available at the top of each hour. The system creates files, writes files, and deletes files in the mount directory of the disk. If the operations fail, the alarm is generated. If the operations succeed, the disk is available, and the alarm is cleared.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12063	Major	Yes

Parameters

Parameter	Description
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the name of the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
DiskName	Specifies the disk for which the alarm is generated.

Impact on the System

Data read or write on the data disk fails, and services are abnormal.

Possible Causes

- The permission of the disk mount directory is abnormal.
- There are disk bad sectors.

Procedure

Check whether the permission of the disk mount directory is normal.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and view the IP address of the host and **DiskName** for the disk for which the alarm is generated.

Step 2 Log in to the host where the alarm is generated as user **root**.

Step 3 Run the **df -h |grep DiskName** command to obtain the mount point and check whether the permission of the mount directory is unwritable or unreadable.

- If it is, go to **Step 4**.
- If it is not, go to **Step 8**.

NOTE

If the permission of the mount directory is 000 or the owner is **root**, the mount directory is unreadable and unwritable.

Step 4 Modify the directory permission.

Step 5 One hour later, check whether this alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 6**.

Step 6 Contact hardware engineers to rectify the disk.

Step 7 One hour later, check whether this alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 8**.

Collect fault information.

Step 8 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 9 Select **NodeAgent** from the **Service** and click **OK**.

Step 10 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.40 ALM-12064 Host Random Port Range Conflicts with Cluster Used Port

Alarm Description

The system checks whether the random port range of the host conflicts with the range of ports used by the Cluster system every hour. The alarm is generated if they conflict. The alarm is automatically cleared when the random port range of the host is changed to the normal range.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12064	Major	Yes

Parameters

Parameter	Description
Source	Specifiestheclusterorsystemforwhich-thealarmisgenerated.
ServiceName	Specifies the name of the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The default port of the Cluster system is occupied. As a result, some processes fail to be started.

Possible Causes

The random port range configuration is modified.

Procedure

Check the random port range of the system.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and view the IP address of the host for which the alarm is generated.

Step 2 Log in to the host where the alarm is generated as user **root**.

Step 3 Run the **cat /proc/sys/net/ipv4/ip_local_port_range** command to obtain the random port range of the host and check whether the minimum value is smaller than 32768.

- If it is, go to **Step 4**.
- If it is not, goto **Step 7**.

Step 4 Run the **vim /etc/sysctl.conf** command to change the value of **net.ipv4.ip_local_port_range** to **32768 61000**. If this parameter does not exist, add the following configuration: **net.ipv4.ip_local_port_range = 32768 61000**.

Step 5 Run the **sysctl -p /etc/sysctl.conf** command for the modification to take effect.

Step 6 One hour later, check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 7**.

Collect fault information.

Step 7 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 8 Select **NodeAgent** for **Service** and click **OK**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.41 ALM-12066 Trust Relationships Between Nodes Become Invalid

Description

The system checks whether the trust relationship between the active OMS node and other Agent nodes is normal every hour. The alarm is generated if the mutual trust fails. This alarm is automatically cleared if this problem is resolved.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12066	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.

Name	Meaning
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

Some operations on the management plane may be abnormal.

Possible Causes

- The **/etc/ssh/sshd_config** configuration file is damaged.
- The password of user **omm** has expired.

Procedure

Check the status of the /etc/ssh/sshd_config configuration file.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm and click ∨ to view the host list in the alarm details.

Step 2 Log in to the active OMS node as user **omm**.

Step 3 Run the **ssh** command, for example, **ssh *host2***, on each node in the alarm details to check whether the connection fails. (***host2*** is a node other than the OMS node in the alarm details.)

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 Open the **/etc/ssh/sshd_config** configuration file on host2 and check whether **AllowUsers** or **DenyUsers** is configured for other nodes.

- If yes, go to **Step 5**.
- If no, contact OS experts.

Step 5 Modify the whitelist or blacklist to ensure that user **omm** is in the whitelist or not in the blacklist. Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Check the status of the password of user omm.

Step 6 Check the interaction information of the **ssh** command.

- If the password of user **omm** is required, go to **Step 7**.
- If message "Enter passphrase for key '/home/omm/.ssh/id_rsa':" is displayed, go to **Step 9**.

Step 7 Check the trust list (**/home/omm/.ssh/authorized_keys**) of user **omm** on the OMS node and host2 node. Check whether the trust list contains the public key file (**/home/omm/.ssh/id_rsa.pub**) of user **omm** on the peer host.

- If yes, contact OS experts.
- If no, add the public key of user **omm** of the peer host to the trust list of the local host.

Step 8 Add the public key of user **omm** of the peer host to the trust list of the local host. Run the **ssh** command, for example, **ssh host2**, on each node in the alarm details to check whether the connection fails. (***host2*** is a node other than the OMS node in the alarm details.)

- If yes, go to **Step 9**.
- If no, check whether the alarm is cleared. If the alarm is cleared, no further action is required; otherwise, go to **Step 9**.

Collect the fault information.

Step 9 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 10 Select **Controller** for **Service** and click **OK**.

Step 11 Click ✎ in the upper right corner to set the log collection time range. Generally, the time range is 10 minutes before and after the alarm generation time. Click **Download**.

Step 12 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

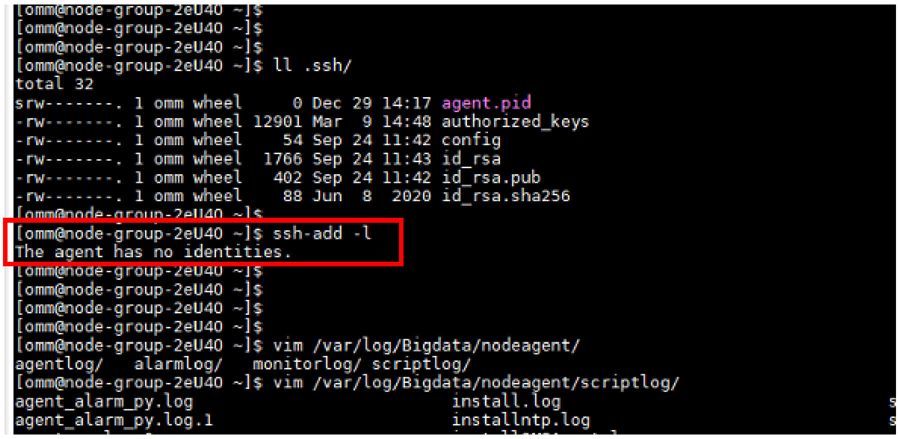
Related Information

Perform the following steps to handle abnormal trust relationships between nodes:

NOTICE

- Perform this operation as user **omm**.
- If the network between nodes is disconnected, rectify the network fault first. Check whether the two nodes are connected to the same security group and whether **hosts.deny** and **hosts.allow** are set.

1. Run the **ssh-add -l** command on both nodes to check whether any identities exist.

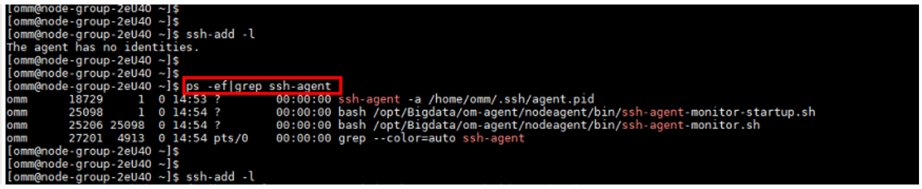


```
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ll .ssh/
total 32
srw-------. 1 omm wheel     0 Dec 29 14:17 agent.pid
-rw-------. 1 omm wheel 12901 Mar  9 14:48 authorized_keys
-rw-------. 1 omm wheel    54 Sep 24 11:42 config
-rw-------. 1 omm wheel  1766 Sep 24 11:43 id_rsa
-rw-------. 1 omm wheel   402 Sep 24 11:42 id_rsa.pub
-rw-------. 1 omm wheel    88 Jun  8  2020 id_rsa.sha256
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh-add -l
The agent has no identities.
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ vim /var/log/Bigdata/nodeagent/
agentlog/   alarmlog/   monitorlog/ scriptlog/
[omm@node-group-2eU40 ~]$ vim /var/log/Bigdata/nodeagent/scriptlog/
agent_alarm_py.log                          install.log                              s
agent_alarm_py.log.1                        installntp.log                           s
```

– If yes, go to **4**.

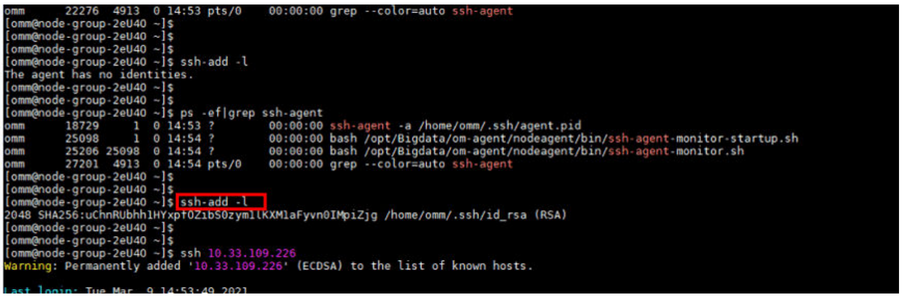
– If no, go to **2**.

2. If no identities are displayed, run the **ps -ef|grep ssh-agent** command to find the **ssh-agent** process, stop the process, and wait for the process to automatically restart.



```
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh-add -l
The agent has no identities.
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ps -ef|grep ssh-agent
omm      18729     1  0 14:53 ?        00:00:00 ssh-agent -a /home/omm/.ssh/agent.pid
omm      25098     1  0 14:54 ?        00:00:00 bash /opt/Bigdata/om-agent/nodeagent/bin/ssh-agent-monitor-startup.sh
omm      25206 25098  0 14:54 ?        00:00:00 bash /opt/Bigdata/om-agent/nodeagent/bin/ssh-agent-monitor.sh
omm      27201  4913  0 14:54 pts/0    00:00:00 grep --color=auto ssh-agent
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh-add -l
```

3. Run the **ssh-add -l** command to check whether the identities have been added. If yes, manually run the **ssh** command to check whether the trust relationship is normal.



```
omm      22276  4913  0 14:53 pts/0    00:00:00 grep --color=auto ssh-agent
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh-add -l
The agent has no identities.
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ps -ef|grep ssh-agent
omm      18729     1  0 14:53 ?        00:00:00 ssh-agent -a /home/omm/.ssh/agent.pid
omm      25098     1  0 14:54 ?        00:00:00 bash /opt/Bigdata/om-agent/nodeagent/bin/ssh-agent-monitor-startup.sh
omm      25206 25098  0 14:54 ?        00:00:00 bash /opt/Bigdata/om-agent/nodeagent/bin/ssh-agent-monitor.sh
omm      27201  4913  0 14:54 pts/0    00:00:00 grep --color=auto ssh-agent
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh-add -l
2048 SHA256:uChnRUbhh1HYxpf0Z1bS0zym1lKXMlaFyvn0IMpiZjg /home/omm/.ssh/id_rsa (RSA)
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh 10.33.109.226
Warning: Permanently added '10.33.109.226' (ECDSA) to the list of known hosts.

Last login: Tue Mar  9 14:53:49 2021
```

4. If identities exist, check whether the **/home/omm/.ssh/authorized_keys** file contains the information in the **/home/omm/.ssh/id_rsa.pub** file of the peer node. If it does not, manually add the information.

5. Check whether the permissions on the files in the **/home/omm/.ssh** directory are modified.

6. Check the **/var/log/Bigdata/nodeagent/scriptlog/ssh-agent-monitor.log** file.

7. If the **/home** directory of user **omm** is deleted, contact MRS support personnel for assistance.

9.13.42 ALM-12067 Tomcat Resource Is Abnormal

Description

HA checks the Tomcat resources of Manager every 85 seconds. This alarm is generated when HA detects that the Tomcat resources are abnormal for two consecutive times.

This alarm is cleared when HA detects that the Tomcat resources become normal.

Resource Type of Tomcat is **Single-active**. Active/standby will be triggered upon resource exceptions. When this alarm is generated, the active/standby switchover is complete and new Tomcat resources have been enabled on the new active Manager. In this case, this alarm is cleared. This alarm is used to notify users of the cause of the active/standby Manager switchover.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12067	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

- The active/standby Manager switchover occurs.
- The Tomcat process repeatedly restarts.

Possible Causes

- The Tomcat directory permission is abnormal, and the Tomcat process is abnormal.

Procedure

Check whether the permission on the Tomcat directory is normal.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and click ⌄ to view the IP address of the host for which the alarm is generated.

Step 2 Log in to the alarm host as user **root**.

Step 3 Run the **su - omm** command to switch to user **omm**.

Step 4 Run the **vi $BIGDATA_LOG_HOME/omm/oms/ha/scriptlog/tomcat.log** command to check whether the Tomcat resource log contains keyword **Cannot find *XXX*** and rectify the file permission based on the keyword.

Step 5 After 5 minutes, check whether the alarm is automatically cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 In the **Services** area, select **OmmServer** and **Tomcat**, and click **OK**.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.43 ALM-12068 ACS Resource Exception

Description

HA checks the ACS resources of Manager every 80 seconds. This alarm is generated when HA detects that the ACS resources are abnormal for two consecutive times.

This alarm is cleared when HA detects that the ACS resources are normal.

Resource Type of ACS is **Single-active**. Active/standby will be triggered upon resource exceptions. When this alarm is generated, the active/standby switchover is complete and new ACS resources have been enabled on the new active Manager. In this case, this alarm is cleared. This alarm is used to notify users of the cause of the active/standby Manager switchover.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12068	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

- The active/standby Manager switchover occurs.
- The ACS process repeatedly restarts, which may cause the FusionInsight Manager login failure.

Possible Causes

The ACS process is abnormal.

Procedure

Check whether the ACS process is normal.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and click ⌄ to view the name of the host for which the alarm is generated.

Step 2 Log in to the alarm host as user **root**.

Step 3 Run the **su - omm** command and then **sh ${BIGDATA_HOME}/om-server/OMS/workspace0/ha/module/hacom/script/status_ha.sh** to check whether the status of the ACS resources managed by the HA is normal. In the single-node system, the ACS resource is in the normal state. In the dual-node system, the ACS resource is in the normal state on the active node and in the stopped state on the standby node.

- If yes, go to **Step 6**.
- If no, go to **Step 4**.

Step 4 Run the **vi $BIGDATA_LOG_HOME/omm/oms/ha/scriptlog/acs.log** command to check whether the ACS resource log of HA contains the keyword **ERROR**. If yes, analyze the logs to locate the resource exception cause and fix the exception.

Step 5 After 5 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 In the **Services** area, select **Controller** and **OmmServer**, and click **OK**.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.44 ALM-12069 AOS Resource Exception

Description

HA checks the AOS resources of Manager every 81 seconds. This alarm is generated when HA detects that the AOS resources are abnormal for two consecutive times.

This alarm is cleared when HA detects that the AOS resources become normal.

Resource Type of AOS is **Single-active**. Active/standby will be triggered upon resource exceptions. When this alarm is generated, the active/standby switchover is complete and new AOS resources have been enabled on the new active Manager. In this case, this alarm is cleared. This alarm is used to notify users of the cause of the active/standby Manager switchover.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12069	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

- The active/standby Manager switchover occurs.
- The AOS process repeatedly restarts, which may cause the FusionInsight Manager login failure.

Possible Causes

The AOS process is abnormal.

Procedure

Check whether the AOS process is normal.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and click ⌄ to view the name of the host for which the alarm is generated.

Step 2 Log in to the alarm host as user **root**.

Step 3 Run the **sh ${BIGDATA_HOME}/om-server/OMS/workspace0/ha/module/hacom/script/status_ha.sh** command to check whether the status of the AOS resources managed by the HA is normal. In the single-node system, the AOS resource is in the normal state. In the dual-node system, the AOS resource is in the normal state on the active node and in the stopped state on the standby node.

- If yes, go to **Step 6**.
- If no, go to **Step 4**.

Step 4 Run the **vi $BIGDATA_LOG_HOME/omm/oms/ha/scriptlog/aos.log** command to check whether the AOS resource log of HA contains the keyword **ERROR**. If yes, analyze the logs to locate the resource exception cause and fix the exception.

Step 5 After 5 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 In the **Services** area, select **Controller** and **OmmServer**, and click **OK**.

Step 8 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.45 ALM-12070 Controller Resource Is Abnormal

Alarm Description

HA checks the controller resources of Manager every 80 seconds. This alarm is generated when HA detects that the controller resources are abnormal for 2 consecutive times.

This alarm is cleared when the Controller resource is normal.

Resource Type of Controller is **Single-active**. Active/standby will be triggered upon resource exceptions. When this alarm is generated, the active/standby switchover is complete and new Controller resources have been enabled on the new active FusionInsight Manager. In this case, this alarm is cleared. This alarm is used to notify users of the cause of the active/standby switchover.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12070	Major	Yes

Parameters

Parameter	Description
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the name of the service for which the alarm is generated.

Parameter	Description
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

- The active/standby FusionInsight Manager switchover occurs.
- The Controller process repeatedly restarts, which may cause the FusionInsight Manager login failure.

Possible Causes

The Controller process is abnormal.

Procedure

Check whether the controller process is normal.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and view the name of the host for which the alarm is generated.

Step 2 Log in to the host for which the alarm is generated as user **root**.

Step 3 Run the **su - omm** command to switch to user **omm**.Run the **sh ${BIGDATA_HOME}/om-server/OMS/workspace0/ha/module/hacom/script/status_ha.sh** command to check whether the status of the Controller resources managed by the HA is normal. In the single-node system, the Controller resource is in the normal state. In the dual-node system, the Controller resource is in the normal state on the active node and in the stopped state on the standby node.

- If it is, go to **Step 6**.
- If it is not, go to **Step 4**.

Step 4 Run the **vi $BIGDATA_LOG_HOME/omm/oms/ha/scriptlog/controller.log** command to view the Controller resource logs, and run the **vi $BIGDATA_LOG_HOME/controller/controller.log** command to view the Controller running logs, check whether the keyword **ERROR** exists. Analyze the logs to locate and rectify the fault.

Step 5 Five minutes later, check whether this alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 6**.

Collect fault information.

Step 6 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 7 Select **Controller** and **OmmServe** for **Service** and click **OK**.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour before and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.46 ALM-12071 Httpd Resource Is Abnormal

Description

HA checks the httpd resources of Manager every 120 seconds. This alarm is generated when HA detects that the httpd resources are abnormal for 10 consecutive times.

This alarm is cleared when the httpd resource is normal.

Resource Type of httpd is **Single-active**. Active/standby will be triggered upon resource exceptions. When this alarm is generated, the active/standby switchover is complete and new httpd resources have been enabled on the new active FusionInsight Manager. In this case, this alarm is cleared. This alarm is used to notify users of the cause of the active/standby switchover.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12071	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.

Name	Meaning
HostName	Specifies the host for which the alarm is generated.

Impact on the System

- The active/standby FusionInsight Manager switchover occurs.
- The httpd process is repeatedly restarts, which may lead to the failure to visit the native service UI.

Possible Causes

The httpd process is abnormal.

Procedure

Check whether the httpd process is abnormal.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and view the name of the host for which the alarm is generated.

Step 2 Log in to the host for which the alarm is generated as user **root**.

Step 3 Run the **su - omm** command to switch to user **omm**.

Step 4 Run the **sh ${BIGDATA_HOME}/om-server/OMS/workspace0/ha/module/hacom/script/status_ha.sh** command to check whether the status of the httpd resources managed by the HA is normal. In the single-node system, the httpd resource is in the normal state. In the dual-node system, the httpd resource is in the normal state on the active node and in the stopped state on the standby node.

- If it is, go to **Step 7**.
- If it is not, go to **Step 5**.

Step 5 Run the **vi $BIGDATA_LOG_HOME/omm/oms/ha/scriptlog/httpd.log** command to view the httpd resource logs, check whether the keyword **ERROR** exists. Analyze the logs to locate and rectify the fault.

Step 6 Five minutes later, check whether this alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 7**.

Collect fault information.

Step 7 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 8 Select **Controller** and **OmmServer** for **Service** and click **OK**.

Step 9 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 1 hour before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.47 ALM-12072 FloatIP Resource Is Abnormal

Description

HA checks the floatip resources of Manager every 9 seconds. This alarm is generated when HA detects that the floatip resources are abnormal for 3 consecutive times.

This alarm is cleared when the FloatIP resource is normal.

Resource Type of FloatIP is **Single-active**. Active/standby will be triggered upon resource exceptions. When this alarm is generated, the active/standby switchover is complete and new FloatIP resources have been enabled on the new active FusionInsight Manager. In this case, this alarm is cleared. This alarm is used to notify users of the cause of the active/standby switchover.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12072	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

- The active/standby FusionInsight Manager switchover occurs.
- The FloatIP process is repeatedly restarts, which may lead to the failure to visit the native service UI.

Possible Causes

- The floating IP address is abnormal.

Procedure

Check the floating IP address status of the active management node.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and view the address of the host for which the alarm is generated and the resource name.

Step 2 Log in to the active management node as user **root**.

Step 3 Run the following command, go to the **${BIGDATA_HOME}/om-server/om/sbin/** directory.

su - omm

cd ${BIGDATA_HOME}/om-server/om/sbin/

Step 4 Run the **sh status-oms.sh** command, and execute the **status-oms.sh** script to check whether the floating IP address of the active FusionInsight Manager is normal. View the command output, locate the row where **ResName** is **floatip**, and check whether the following information is displayed.

For example:

```
10-10-10-160 floatip Normal Normal Single_active
```

- If it is, go to **Step 8**.
- If it is not, go to **Step 5**.

Step 5 Run the **ifconfig** command to check whether the NIC with the floating IP address exists.

- If it does, go to **Step 8**.
- If it does not, go to **Step 6**.

Step 6 Run the **ifconfig** *NIC name Floating IPaddress* netmask *Subnet mask* command to reconfigure the NIC with the floating IP address. (For example, **ifconfig eth0 10.10.10.102 netmask 255.255.255.0**).

Step 7 Five minutes later, check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 8**.

Collect fault information.

Step 8 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 9 Select **Controller** and **OmmServer** for **Service** and click **OK**.

Step 10 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 1 hour before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 11 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.48 ALM-12073 CEP Resource Is Abnormal

Description

HA checks the cep resources of Manager every 60 seconds. This alarm is generated when HA detects that the cep resources are abnormal for 2 consecutive times.

This alarm is cleared when the CEP resource is normal.

Resource Type of CEP is **Single-active**. Active/standby will be triggered upon resource exceptions. When this alarm is generated, the active/standby switchover is complete and new CEP resources have been enabled on the new active FusionInsight Manager. In this case, this alarm is cleared. This alarm is used to notify users of the cause of the active/standby switchover.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12073	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.

Name	Meaning
HostName	Specifies the host for which the alarm is generated.

Impact on the System

- The active/standby FusionInsight Manager switchover occurs.
- The CEP process repeatedly restarts, causing monitoring data to be abnormal.

Possible Causes

The CEP process is abnormal.

Procedure

Check whether the CEP process is abnormal.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and view the name of the host for which the alarm is generated.

Step 2 Log in to the host for which the alarm is generated as user **root**.

Step 3 Run the **su -omm** command and then the **sh ${BIGDATA_HOME}/om-server/OMS/workspace0/ha/module/hacom/script/status_ha.sh** command to check whether the status of the CEP resources managed by the HA is normal. In the single-node system, the CEP resource is in the normal state. In the dual-node system, the CEP resource is in the normal state on the active node and in the stopped state on the standby node.

- If it is, go to **Step 6**.
- If it is not, go to **Step 4**.

Step 4 Run the **vi $BIGDATA_LOG_HOME/omm/oms/cep/cep.log** and **vi $BIGDATA_LOG_HOME/omm/oms/cep/scriptlog/cep_ha.log** commands to view the CEP resource logs, check whether the keyword **ERROR** exists. Analyze the logs to locate and rectify the fault.

Step 5 Five minutes later, check whether this alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 6**.

Collect fault information.

Step 6 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 7 Select **Controller** and **OmmServer** for **Service** and click **OK**.

Step 8 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 1 hour before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.49 ALM-12074 FMS Resource Is Abnormal

Description

HA checks the fms resources of Manager every 60 seconds. This alarm is generated when HA detects that the fms resources are abnormal for 2 consecutive times.

This alarm is cleared when the FMS resource is normal.

Resource Type of FMS is **Single-active**. Active/standby will be triggered upon resource exceptions. When this alarm is generated, the active/standby switchover is complete and new FMS resources have been enabled on the new active FusionInsight Manager. In this case, this alarm is cleared. This alarm is used to notify users of the cause of the active/standby switchover.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12074	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

- The active/standby FusionInsight Manager switchover occurs.
- The FMS process repeatedly restarts. As a result, alarm information may fail to be reported.

Possible Causes

The FMS process is abnormal.

Procedure

Check whether the FMS process is abnormal.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and view the name of the host for which the alarm is generated.

Step 2 Log in to the host for which the alarm is generated as user **root**.

Step 3 Run the **su -omm** command and then the **sh ${BIGDATA_HOME}/om-server/OMS/workspace0/ha/module/hacom/script/status_ha.sh** command to check whether the status of the FMS resources managed by the HA is normal. In the single-node system, the FMS resource is in the normal state. In the dual-node system, the FMS resource is in the normal state on the active node and in the stopped state on the standby node.

- If it is, go to **Step 6**.
- If it is not, go to **Step 4**.

Step 4 Run the **vi $BIGDATA_LOG_HOME/omm/oms/fms/fms.log** and **vi $BIGDATA_LOG_HOME/omm/oms/fms/scriptlog/fms_ha.log** commands to view the FMS resource logs, check whether the keyword **ERROR** exists. Analyze the logs to locate and rectify the fault.

Step 5 5 minutes later, check whether this alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 6**.

Collect fault information.

Step 6 On FusionInsight Manager, choose **O&M**> **Log** > **Download**.

Step 7 Select **Controller** and **OmmServer** for **Service** and click **OK**.

Step 8 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 1 hour before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.50 ALM-12075 PMS Resource Is Abnormal

Description

HA checks the pms resources of Manager every 55 seconds. This alarm is generated when HA detects that the pms resources are abnormal for three consecutive times.

This alarm is cleared when the PMS resource is normal.

Resource Type of PMS is **Single-active**. Active/standby will be triggered upon resource exceptions. When this alarm is generated, the active/standby switchover is complete and new PMS resources have been enabled on the new active FusionInsight Manager. In this case, this alarm is cleared. This alarm is used to notify users of the cause of the active/standby switchover.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12075	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

- The active/standby FusionInsight Manager switchover occurs.
- The PMS process repeatedly restarts, causing monitoring information to be abnormal.

Possible Causes

The PMS process is abnormal.

Procedure

Check whether the PMS process is abnormal.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and view the name of the host for which the alarm is generated.

Step 2 Log in to the host for which the alarm is generated as user **root**.

Step 3 Run the **su -omm** command and then the **sh ${BIGDATA_HOME}/om-server/OMS/workspace0/ha/module/hacom/script/status_ha.sh** command to check whether the status of the PMS resources managed by the HA is normal. In the single-node system, the PMS resource is in the normal state. In the dual-node system, the PMS resource is in the normal state on the active node and in the stopped state on the standby node.

- If it is, go to **Step 6**.
- If it is not, go to **Step 4**.

Step 4 Run the **vi $BIGDATA_LOG_HOME/omm/oms/pms/pms.log** and **vi $BIGDATA_LOG_HOME/omm/oms/pms/scriptlog/pms_ha.log** commands to view the PMS resource logs, check whether the keyword **ERROR** exists. Analyze the logs to locate and rectify the fault.

Step 5 Five minutes later, check whether this alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 6**.

Collect fault information.

Step 6 On FusionInsight Manager, choose **O&M**> **Log** > **Download**.

Step 7 Select **Controller** and **OmmServer** for **Service** and click **OK**.

Step 8 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 1 hour before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.51 ALM-12076 GaussDB Resource Is Abnormal

Description

HA checks the Manager database every 10 seconds. This alarm is generated when HA detects that the database is abnormal for 3 consecutive times.

This alarm is cleared when the database is normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12076	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

If databases are abnormal, all core services and related service processes, such as alarms and monitoring functions, are affected.

Possible Causes

An exception occurs in the database.

Procedure

Check the database status of the active and standby management nodes.

Step 1 Log in to the active and standby management nodes respectively as user **root**. Run the **su - ommdba** command to switch to user **ommdba**, and then run the **gs_ctl query** command to check whether the following information is displayed in the command output.

Command output of the active management node:

```
Ha state:
      LOCAL_ROLE: Primary
      STATIC_CONNECTIONS           : 1
      DB_STATE                    : Normal
      DETAIL_INFORMATION           : user/password invalid
Senders info:
      No information
Receiver info:
      No information
```

Command output of the standby management node:

```
Ha state:
      LOCAL_ROLE: Standby
```

```
        STATIC_CONNECTIONS            : 1
        DB_STATE                    : Normal
        DETAIL_INFORMATION            : user/password invalid
Senders info:
        No information
Receiver info:
        No information
```

- If it is, go to **Step 3**.
- If it is not, go to **Step 2**.

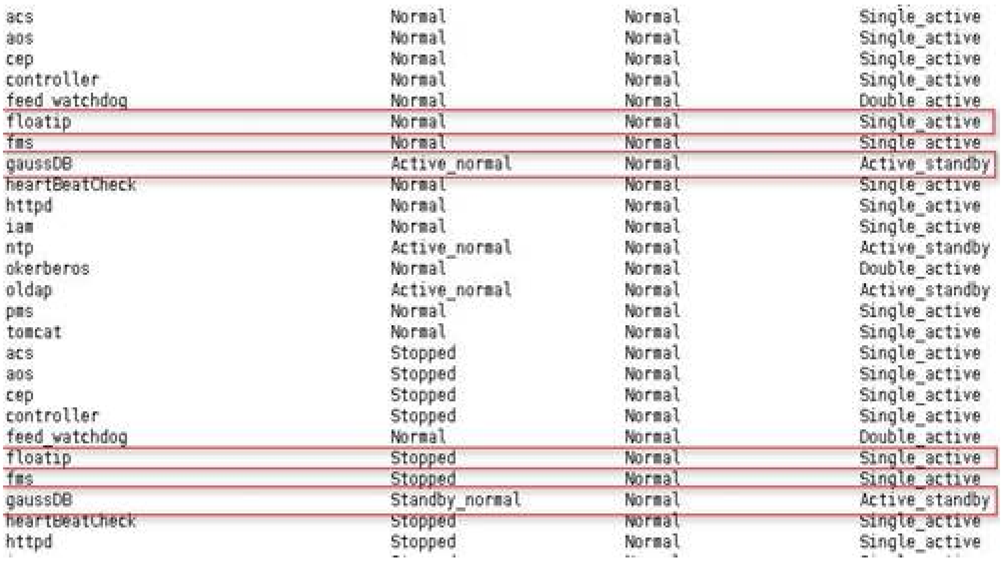
Step 2 Contact the network administrator to check whether the network is faulty.

- If it is, go to **Step 3**.
- If it is not, go to **Step 5**.

Step 3 Five minutes later, check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 4**.

Step 4 Log in to the active and standby management nodes, run the **su -omm** command to switch to user **omm**, go to the **${BIGDATA_HOME} /om-server/om/sbin/** directory, and run the **status-oms.sh** script to check whether the floating IP addresses and GaussDB resources of the active and standby FusionInsight Managers are in the status shown in the following figure.



```
acs                   Normal            Normal            Single_active
aos                   Normal            Normal            Single_active
cep                   Normal            Normal            Single_active
controller            Normal            Normal            Single_active
feed_watchdog         Normal            Normal            Double_active
floatip               Normal            Normal            Single_active
fms                   Normal            Normal            Single_active
gaussDB               Active_normal     Normal            Active_standby
heartBeatCheck        Normal            Normal            Single_active
httpd                 Normal            Normal            Single_active
iam                   Normal            Normal            Single_active
ntp                   Active_normal     Normal            Active_standby
okerberos             Normal            Normal            Double_active
oldap                 Active_normal     Normal            Active_standby
pms                   Normal            Normal            Single_active
tomcat                Normal            Normal            Single_active
acs                   Stopped           Normal            Single_active
aos                   Stopped           Normal            Single_active
cep                   Stopped           Normal            Single_active
controller            Stopped           Normal            Single_active
feed_watchdog         Normal            Normal            Double_active
floatip               Stopped           Normal            Single_active
fms                   Stopped           Normal            Single_active
gaussDB               Standby_normal    Normal            Active_standby
heartBeatCheck        Stopped           Normal            Single_active
httpd                 Stopped           Normal            Single_active
```

- If they are, find the alarm in the alarm list and manually clear the alarm.
- If they are not, go to **Step 5**.

Collect fault information.

Step 5 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 6 Select **OmmServer** for **Service** and click **OK**.

Step 7 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 10 minutes before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 8 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.52 ALM-12077 User omm Expired

Description

The system starts at 00:00 every day to check whether user **omm** has expired every eight hours. This alarm is generated if the user account has expired.

This alarm is cleared when the expiration time of user **omm** is changed and the user account status becomes normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12077	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

User **omm** has expired. The node trust relationship is unavailable, and FusionInsight Manager cannot manage the services.

Possible Causes

User **omm** has expired.

Procedure

Check whether user omm in the system has expired.

Step 1 Log in to the faulty node as user **root**.

Run the **chage -l omm**command to view the information about the password of user **omm**.

Step 2 View the value of **Account expires** to check whether the user configurations have expired.

NOTE

If the parameter value is **never**, the user configurations never expire.

- If they do, go to **Step 3**.
- If they do not, go to **Step 4**.

Step 3 Run the **chage -E 'yyyy-MM-dd' omm** command to set the expiration time of user **omm**. Eight hours later, check whether the alarm is automatically cleared.

- If it is, no further action is required.
- If it is not, go to **Step 4**.

Collect fault information.

Step 4 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 5 Select **NodeAgent** for **Service** and click **OK**.

Step 6 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 10 minutes before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 7 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.53 ALM-12078 Password of User omm Expired

Description

The system starts at 00:00 every day to check whether the password of user **omm** has expired every 8 hours. This alarm is generated if the password has expired.

This alarm is cleared when the expiration time of user **omm** password is changed and the user password status becomes normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12078	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The password of user **omm** has expired. The node trust relationship is unavailable, and FusionInsight Manager cannot manage the services.

Possible Causes

The password of user **omm** has expired.

Procedure

Check whether the password of user omm in the system has expired.

Step 1 Log in to the faulty node as user **root**.

Run the **chage -l omm**command to view the information about the password of user **omm**.

Step 2 View the value of **Password expires** to check whether the user configurations have expired.

NOTE

If the parameter value is **never**, the user configurations never expire.

- If they do, go to **Step 3**.
- If they do not, go to **Step 4**.

Step 3 Run the **chage -M '*days*' omm** command to set the validity period of the password for user **omm**. Eight hours later, check whether the alarm is automatically cleared.

- If it is, no further action is required.
- If it is not, go to **Step 4**.

Collect fault information.

Step 4 On FusionInsight Manager, choose **O&M**> **Log** > **Download**.

Step 5 Select **NodeAgent** for **Service** and click **OK**.

Step 6 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 10 minutes before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 7 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.54 ALM-12079 User omm Is About to Expire

Description

The system starts at 00:00 every day to check whether user **omm** is about to expire every 8 hours. This alarm is generated if the user account will expire no less than 15 days later.

This alarm is cleared when the expiration time of user **omm** is changed and the user account status becomes normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12079	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

User **omm** has expired. The node trust relationship is unavailable, and FusionInsight Manager cannot manage the services.

Possible Causes

The account of user **omm** is about to expire.

Procedure

Check whether user omm is about to expire.

Step 1 Log in to the faulty node as user **root**.

Run the **chage -l omm** command to view the information about the password of user **omm**.

Step 2 View the value of **Account expires** to check whether the user configurations are about to expire.

NOTE

If the parameter value is **never**, the user and password are valid permanently; if the value is a date, check whether the user and password are about to expire within 15 days.

- If they are, go to **Step 3**.
- If they are not, go to **Step 4**.

Step 3 Run the **chage -E** *'yyyy-MM-dd'* **omm** command to set the validity period of user **omm**. Eight hours later, check whether the alarm is automatically cleared.

- If it is, no further action is required.
- If it is not, go to **Step 4**.

Collect fault information.

Step 4 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 5 Select **NodeAgent** for **Service** and click **OK**.

Step 6 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 10 minutes before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 7 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.55 ALM-12080 Password of User omm Is About to Expire

Description

The system starts at 00:00 every day to check whether the password of user **omm** is about to expire every 8 hours. This alarm is generated if the password will expire no less than 15 days later.

This alarm is cleared when the expiration time of user **omm** password is reset and the user password status becomes normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12080	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The password of user **omm** has expired. The node trust relationship is unavailable, and FusionInsight Manager cannot manage the services.

Possible Causes

The password of user **omm** is about to expire.

Procedure

Check whether the password of user omm in the system is about to expire.

Step 1 Log in to the faulty node as user **root**.

Run the **chage -l omm**command to view the information about the password of user **omm**.

Step 2 View the value of **Password expires** to check whether the user configurations are about to expire.

NOTE

If the parameter value is **never**, the user and password are valid permanently; if the value is a date, check whether the user and password are about to expire within 15 days.

- If they are, go to **Step 3**.
- If they are not, go to **Step 4**.

Step 3 Run the **chage -M '***days***' omm** command to set the validity period of the password for user **omm**. Eight hours later, check whether the alarm is automatically cleared.

- If it is, no further action is required.
- If it is not, go to **Step 4**.

Collect fault information.

Step 4 On FusionInsight Manager, choose **O&M**> **Log** > **Download**.

Step 5 Select **NodeAgent** for **Service** and click **OK**.

Step 6 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 10 minutes before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 7 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.56 ALM-12081 User ommdba Expired

Description

The system starts at 00:00 every day to check whether user **ommdba** has expired every 8 hours. This alarm is generated if the user account has expired.

This alarm is cleared when the expiration time of user **ommdba** is reset and the user account status becomes normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12081	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The OMS database cannot be managed and data cannot be accessed.

Possible Causes

The account of user **ommdba** for the host has expired.

Procedure

Check whether user ommdba has expired.

Step 1 Log in to the faulty node as user **root**.

Run the **chage -l ommdba** command to view the information about the password of user **ommdba**.

Step 2 View the value of **Account expires** to check whether the user configurations have expired.

NOTE

If the parameter value is **never**, the user and password are valid permanently; if the value is a date, check whether the user and password have expired.

- If they do, go to **Step 3**.
- If they do not, go to **Step 4**.

Step 3 Run the **chage -E** *'yyyy-MM-dd'* **omm** command to set the validity period of user **ommdba**. Eight hours later, check whether the alarm is automatically cleared.

- If it is, no further action is required.
- If it is not, go to **Step 4**.

Collect fault information.

Step 4 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 5 Select **NodeAgent** for **Service** and click **OK**.

Step 6 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 10 minutes before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 7 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.57 ALM-12082 User ommdba Is About to Expire

Description

The system starts at 00:00 every day to check whether user **ommdba** is about to expire every 8 hours. This alarm is generated if the user account will expire no less than 15 days later.

This alarm is cleared when the expiration time of user **ommdba** is reset and the user account status becomes normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12082	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The OMS database cannot be managed and data cannot be accessed.

Possible Causes

The account of user **ommdba** for the host is about to expire.

Procedure

Check whether user ommdba is about to expire.

Step 1 Log in to the faulty node as user **root**.

Run the **chage -l ommdba** command to view the information about user **ommdba**.

Step 2 View the value of **Account expires** to check whether the user configurations are about to expire.

NOTE

If the parameter value is **never**, the user and password are valid permanently; if the value is a date, check whether the user and password are about to expire within 15 days.

- If they are, go to **Step 3**.
- If they are not, go to **Step 4**.

Step 3 Run the **chage -E** *'yyyy-MM-dd'* **ommdba** command to set the validity period of user **ommdba**. Eight hours later, check whether the alarm is automatically cleared.

- If it is, no further action is required.
- If it is not, go to **Step 4**.

Collect fault information.

Step 4 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 5 Select **NodeAgent** for **Service** and click **OK**.

Step 6 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 10 minutes before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 7 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.58 ALM-12083 Password of User ommdba Is About to Expire

Description

The system starts at 00:00 every day to check whether the password of user **ommdba** is about to expire every 8 hours. This alarm is generated if the password is about to expire no less than 15 days later.

This alarm is cleared when the expiration time of user **ommdba** password is reset and the user password status becomes normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12083	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The OMS database cannot be managed and data cannot be accessed.

Possible Causes

The password of user **ommdba** is about to expire.

Procedure

Check whether the password of user ommdba in the system is about to expire.

Step 1 Log in to the faulty node as user **root**.

Run the **chage -l ommdba** command to view the information about the password of user **ommdba**.

Step 2 View the value of **Password expires** to check whether the user configurations are about to expire.

NOTE

If the parameter value is **never**, the user and password are valid permanently; if the value is a date, check whether the user and password are about to expire within 15 days.

- If they are, go to **Step 3**.
- If they are not, go to **Step 4**.

Step 3 Run the **chage -M** *'days'* **ommdba** command to set the validity period of the password for user **ommdba**. Eight hours later, check whether the alarm is automatically cleared.

- If it is, no further action is required.
- If it is not, go to **Step 4**.

Collect fault information.

Step 4 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 5 Select **NodeAgent** for **Service** and click **OK**.

Step 6 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 10 minutes before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 7 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.59 ALM-12084 Password of User ommdba Expired

Description

The system starts at 00:00 every day to check whether the password of user **ommdba** has expired every 8 hours. This alarm is generated if the password has expired.

This alarm is cleared when the expiration time of user **ommdba** password is reset and the user password status becomes normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12084	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The password of user **ommdba** has expired. The node trust relationship is unavailable, and FusionInsight Manager cannot manage the services.

Possible Causes

The password of user **ommdba** for the host has expired.

Procedure

Check whether the password of user ommdba in the system has expired.

Step 1 Log in to the faulty node as user **root**.

Run the **chage -l ommdba** command to view the information about the password of user **ommdba**.

Step 2 View the value of **Password expires** to check whether the user configurations have expired.

NOTE

If the parameter value is **never**, the user and password are valid permanently; if the value is a date, check whether the user and password have expired.

- If they do, go to **Step 3**.
- If they do not, go to **Step 4**.

Step 3 Run the **chage -M** *'days'* **ommdba** command to set the validity period of the password for user **ommdba**. Eight hours later, check whether the alarm is automatically cleared.

- If it is, no further action is required.
- If it is not, go to **Step 4**.

Collect fault information.

Step 4 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 5 Select **NodeAgent** for **Service** and click **OK**.

Step 6 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 10 minutes before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 7 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.60 ALM-12085 Service Audit Log Dump Failure

Description

The system dumps service audit logs at 03:00 every day and stores them on the OMS node. This alarm is generated when the dump fails. This alarm is cleared when the next dump succeeds.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12085	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The service audit logs may be lost.

Possible Causes

- The service audit logs are oversized.
- The OMS backup storage space is insufficient.
- The storage space of a host where the service is located is insufficient.

Procedure

Check whether the service audit logs are oversized.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and view the IP address of the host and additional information for which the alarm is generated.

Step 2 Log in to the host where the alarm is generated as user **root**.

Step 3 Run the **vi ${BIGDATA_LOG_HOME}/controller/scriptlog/getLogs.log** command to check whether the keyword "LOG SIZE is more than 5000MB" can be searched.

- If it can, go to **Step 4**.
- If it cannot, go to **Step 5**.

Step 4 Check whether the oversized service audit logs are caused by exceptions.

The OMS backup storage space is insufficient.

Step 5 Run the **vi ${BIGDATA_LOG_HOME}/controller/scriptlog/getLogs.log** command to check whether the keyword "Collect log failed, too many logs on" can be searched.

- If it can, obtain the host IP address following the keyword "Collect log failed, too many logs on", and go to **Step 6**.
- If it cannot, go to **Step 11**.

Step 6 Log in to the host with the IP address obtained in **Step 5** as user **root**.

Step 7 Run the **vi {BIGDATA_LOG_HOME}/nodeagent/scriptlog/collectLog.log** command to check whether the keyword "log size exceeds" can be searched.

- If it can, go to **Step 9**.
- If it cannot, go to **Step 8**.

Step 8 Check whether the alarm additional information contains the keyword "no enough space".

- If yes, go to **Step 9**.
- If no, go to**Step 11**.

Step 9 Perform the following operations to expand the disk capacity or reduce the maximum number of audit log backups:

- Expand the capacity of the OMS node.
- Run the following command to edit the file and decrease the value of **MAX_NUM_BK_AUDITLOG**.

  vi ${CONTROLLER_HOME}/etc/om/componentsauditlog.properties

Step 10 In the next execution period, 03:00, check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 11**.

Check whether the space of the host where the service is located is insufficient.

Step 11 Run the **vi ${BIGDATA_LOG_HOME}/controller/scriptlog/getLogs.log** command to check whether the keyword "Collect log failed, no enough space on *hostIp*" can be searched.

- If it can, obtain the IP address of the abnormal host and go to **Step 12**.
- If it cannot, go to **Step 15**.

Step 12 Log in to the host with the IP address obtained as user **root**, and run the **df "$BIGDATA_HOME/tmp" -lP | tail -1 | awk '{print ($4/1024)}'** command to obtain the remaining space of the host log directory. Check whether the value is less than 1000 MB.

- If it is, go to **Step 13**.
- If it is not, go to **Step 15**.

Step 13 Expand the capacity of the node

Step 14 In the next execution period, 03:00, check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 15**.

Collect fault information.

Step 15 On FusionInsight Manager, choose **O&M**> **Log** > **Download**.

Step 16 Select **Controller** for **Service** and click **OK**.

Step 17 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 10 minutes before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 18 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.61 ALM-12087 System Is in the Upgrade Observation Period

Description

The system checks whether it is in the upgrade observation period at 00:00 every day and checks whether the duration that it has been in the upgrade observation state exceeds the preset upgrade observation period, 10 days by default. This alarm is generated when the system is in the upgrade observation period and the duration that the system has been in the upgrade observation state exceeds the preset period (10 days by default). This alarm is automatically cleared if the system exits the upgrade observation period after the user performs a rollback or submission.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12087	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Upgrade Observation Period (Days)	Specifies the days that the system is in the upgrade observation period.

Impact on the System

The next upgrade or patch installation will fail.

Possible Causes

The upgrade task is not submitted a specified period of time (10 days by default) after the system upgrade.

Procedure

Check whether the system is in the upgrade observation period.

Step 1 Log in to the active management node as user **root**.

Step 2 Run the following commands to switch to user **omm** and log in to the **omm** database:

su - omm

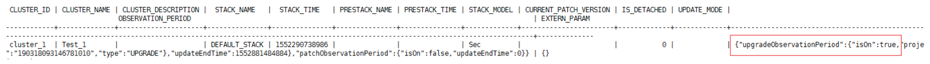
gsql -U omm -W *omm database password* **-p 20015**

Step 3 Run the **select * from OM_CLUSTERS** command to view cluster information.

Step 4 Check whether the value of **upgradObservationPeriod isON** is **true**, as shown in **Figure 9-25**.

- If it is, the system is in the upgrade observation period. Use the UpdateTool to submit the upgrade task. For details, see the upgrade guide of the corresponding version.
- If it is not, go to **Step 6**.

Figure 9-25 Cluster information



```
 CLUSTER_ID | CLUSTER_NAME | CLUSTER_DESCRIPTION |  STACK_NAME   |  STACK_TIME   | PRESTACK_NAME | PRESTACK_TIME | STACK_MODEL | CURRENT_PATCH_VERSION | IS_DETACHED | UPDATE_MODE |
                           OBSERVATION_PERIOD                                                                    | EXTERN_PARAM
------------+--------------+---------------------+---------------+---------------+---------------+---------------+-------------+-----------------------+-------------+-------------+-----------------------------------------
-----------------------------------------------------------------------------------------------------------------+-------------
 cluster_1  | Test_1       |                     | DEFAULT_STACK | 1552290738986 |               |               | Sec         |                       |           0 |             | {"upgradeObservationPeriod":{"isOn":true,"proje
":"190318093146781010","type":"UPGRADE"},"updateEndTime":1552881484884},"patchObservationPeriod":{"isOn":false,"updateEndTime":0}} | {}
```

Step 5 In the early morning of the next day, check whether this alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 7 Select **Controller** from the **Service** and click **OK**.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.62 ALM-12089 Inter-Node Network Is Abnormal

Description

The alarm module checks the network health status of nodes in the cluster every 10 seconds. This alarm is generated when the network between two nodes is unreachable or the network status is unstable.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12089	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

Functions of some components, such as HDFS and ZooKeeper, are affected.

Possible Causes

- The node breaks down.
- The network is faulty.

Procedure

Check the network health status.

Step 1 In the alarm list on FusionInsight Manager, click the drop-down button of the alarm and view **Additional Information**. Record the source IP address and destination IP address of the node for which the alarm is reported.

Step 2 Log in to the node for which the alarm is reported. On the node, ping the target node to check whether the network between the two nodes is normal.

- If yes, go to **6**.
- If no, go to **3**.

Check the node status.

Step 3 On FusionInsight Manager, click **Host** and check whether the host list contains the faulty node to determine whether the faulty node has been removed from the cluster.

- If yes, go to **5**.
- If no, go to **4**.

Step 4 Check whether the faulty node is powered off.

- If yes, start the faulty node and go to **Step 2**.
- If no, contact related personnel to find root cause, if need to remove the faulty nodes from the cluster and go to **5**, otherwise go to **6**.

Step 5 Remove the file **$NODE_AGENT_HOME/etc/agent/hosts.ini** of all nodes in the cluster, and clean up the file **/var/log/Bigdata/unreachable/unreachable_ip_info.log**, and then manually clear the alarm.

Step 6 Wait for 30 seconds and checking if the alarm was been cleared.

- If yes, no further action is required.
- If no, go to **7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **OmmAgent** from the **Service** and click **OK**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.63 ALM-12101 AZ Unhealthy

Description

After the AZ DR function is enabled, the system checks the AZ health status every 5 minutes. This alarm is generated when the system detects that the AZ is subhealthy or unhealthy. This alarm is cleared when the AZ becomes healthy.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12101	Major	Yes

Parameters

Parameter	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
AZName	Specifies the AZ for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The health status of an AZ is determined by whether the health status of storage resources (HDFS), computing resources (Yarn), and key roles in the AZ exceeds the configured threshold.

An AZ is subhealthy when:

- The computing resources (Yarn) are unhealthy, but the storage resources (HDFS) are healthy. Tasks cannot be submitted to the local AZ, but data can still be read and written in the local AZ.
- The computing resources (Yarn) are healthy, but some storage resources (HDFS) are unhealthy. Tasks can be submitted to the local AZ, and some data can be read and written in the local AZ. This depends on the locality of data detected by Spark/Hive scheduling.

An AZ is unhealthy when:

- The computing resources (Yarn) are healthy, but the storage resources (HDFS) are unhealthy. Although tasks can be submitted to the local AZ, data cannot be read or written in the local AZ. As a result, the tasks submitted to the local AZ are invalid.

- The computing resources (Yarn) and storage resources (HDFS) are unhealthy. Tasks cannot be submitted to the local AZ, and data cannot be read or written in the local AZ.
- The health status of key roles except Yarn and HDFS is lower than the configured threshold.

Possible Causes

- The computing resources (Yarn) are unhealthy.
- The storage resources (HDFS) are unhealthy.
- Some storage resources (HDFS) are unhealthy.
- Key roles except Yarn and HDFS are unhealthy.

Procedure

Disable the DR drill.

Step 1 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Cross-AZ HA**. The Cross-AZ HA page is displayed.

Step 2 In the AZ DR list, check whether **Perform DR Drill** in the **Operation** column of the AZ whose health status is **Unhealthy** is gray.

- If yes, go to **Step 4**.
- If no, go to **Step 3**.

Step 3 Click **Restore** in the **Operation** column of the target AZ. Wait 2 minutes and refresh the page to view the health status of the AZ. Check whether the health status is normal.

- If yes, no further action is required.
- If no, go to **Step 4**.

Collect the fault information.

Step 4 Log in to the active management node as user **root**.

Step 5 View logs of unhealthy services.

- HDFS log files are stored in **/var/log/Bigdata/hdfs/nn/hdfs-az-state.log**.
- Yarn log files are stored in **/var/log/Bigdata/yarn/rm/yarn-az-state.log**.
- For other services, view the service health check logs in the corresponding service log directory.

Step 6 Contact O&M personnel and provide detailed log file information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.64 ALM-12102 AZ HA Component Is Not Deployed Based on DR Requirements

Description

The alarm module checks the deployment status of AZ HA components every 5 minutes. This alarm is generated when the components that support DR are not deployed based on DR requirements after AZ is enabled. This alarm is cleared when the components are deployed based on DR requirements.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12102	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.

Impact on the System

The cross-AZ HA capability of a single cluster is affected.

Possible Causes

The roles of the components that support DR are not deployed based on DR requirements.

Procedure

Obtain alarm information.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**.

Step 2 In the alarm list, click ⌄ in the row that contains the alarm and view the roles that are not deployed based on DR requirements in **Additional Information**.

Redeploy the role instance.

Step 3 Choose **Cluster** > **Services** > *Name of the desired service* > **Instance**. On the instance page, redeploy or adjust the role instance.

Step 4 Check whether the alarm is cleared 10 minutes later.

- If yes, no further action is required.
- If no, contact O&M personnel.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.65 ALM-12103 Executor Resource Exception

Description

HA checks the Executor resources of Manager every 30 seconds. This alarm is generated when HA detects that the Executor resources are abnormal for two consecutive times.

This alarm is cleared when the Executor resources are normal.

Resource Type of Executor is **Single-active**. Active/standby will be triggered upon resource exceptions. When this alarm is generated, the active/standby switchover is complete and new Executor resources have been enabled on the new active Manager. In this case, this alarm is cleared. This alarm is used to notify users of the cause of the active/standby Manager switchover.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12103	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

- The active/standby Manager switchover occurs.
- The Executor process keeps restarting. As a result, the cluster page may fail to be accessed.

Possible Causes

The Executor process is abnormal.

Procedure

Check whether the Executor process is abnormal.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and click ⌄ to view the name of the host for which the alarm is generated.

Step 2 Log in to the host for which the alarm is generated as user **root**.

Step 3 Run the **su - omm** command to switch to user **omm**.

Step 4 Run the **sh ${BIGDATA_HOME}/om-server/OMS/workspace0/ha/module/hacom/script/status_ha.sh** command to check whether the status of the Executor resources managed by the HA is normal. In the single-node system, the Executor resource is in the normal state. In the dual-node system, the Executor resource is in the normal state on the active node and in the stopped state on the standby node.

- If yes, go to **Step 7**.
- If no, go to **Step 5**.

Step 5 Run the **vi $BIGDATA_LOG_HOME/omm/oms/ha/scriptlog/executor.log** command to check whether the Executor resource log of HA contains the keyword **ERROR**. If yes, analyze the log to locate the resource exception cause and fix the exception.

Step 6 After 5 minutes, check whether this alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect the fault information.

Step 7 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 8 In the **Services** area, select **Controller** and **OmmServer**, and click **OK**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact O&M personnel and provide the collected logs.

----End

9.13.66 ALM-12104 Abnormal Knox Resources

Description

HA checks the Knox resources of Manager every 70 seconds. This alarm is generated when HA detects that the Knox resources are abnormal for three consecutive times.

This alarm is cleared when HA detects that the Knox resources are normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12104	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

Requests sent by upper-layer services by using Knox cannot be properly processed.

Possible Causes

The Knox process is abnormal.

Procedure

Check whether the Knox process is normal.

Step 1 Log in to FusionInsight Manager. In the alarm list, locate the row that contains the alarm and view the name of the host for which the alarm is generated.

Step 2 Use PuTTY to log in to the host for which the alarm is generated as user **root**.

Step 3 Run the **su - omm** command to switch to user **omm**.

Step 4 Run the **sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh** command to check whether the status of the Knox resources managed by HA is normal. If the status is normal, the Knox resources are normal. Otherwise, the Knox resources are abnormal.

- If yes, go to **Step 7**.
- If no, go to **Step 5**.

Step 5 Run the **vi $BIGDATA_LOG_HOME/omm/oms/ha/scriptlog/knox.log** command to check whether the Knox resource log of HA contains the keyword **ERROR**. If yes, analyze the log to locate the resource exception cause and fix the exception.

Step 6 After 5 minutes, check whether this alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect the fault information.

Step 7 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 8 In the **Services** area, select **Controller** and **OmmServer**, and click **OK**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact O&M personnel and provide the collected logs.

----End

9.13.67 ALM-12110 Failed to get ECS temporary AK/SK

Description

The meta service periodically obtains the temporary AK/SK of the ECS. This alarm is generated when the meta service fails to obtain the temporary AK/SK.

Attribute

Alarm ID	Alarm Severity	Auto Clear
12110	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

In storage-compute decoupling scenarios, the cluster cannot obtain the latest temporary AK/SK, which may lead to failure access to OBS.

Possible Causes

- The meta role of the MRS cluster is abnormal.
- The cluster has been bound to an agency and accessed OBS but has been unbound from the agency. As a result, the cluster has not been bound to any agency.

Procedure

Check the status of the meta role.

Step 1 On FusionInsight Manager of the cluster, choose **O&M** > **Alarm** > **Alarms**. On the page that is displayed, click ⌄ in the row containing the alarm, and determine the IP address of the host for which the alarm is generated.

Step 2 On FusionInsight Manager of the cluster, choose **Cluster** > **Services** > **Meta**. On the page that is displayed, click the **Instance** tab, and check whether the meta role corresponding to the host for which the alarm is generated is normal.

- If yes, go to **Step 4**.
- If no, go to **Step 3**.

Step 3 Select the abnormal role and choose **More** > **Restart Instance** to restart the abnormal meta role. After the restart is complete, check whether the alarm is cleared several minutes later.

- If yes, no further action is required.
- If no, go to **Step 4**.

Rebind the cluster to an agency.

Step 4 Log in to the MRS management console.

Step 5 In the navigation pane on the left, choose **Clusters** > **Active Clusters**. On the page that is displayed, click the cluster name to go to its overview page. Then, check whether the cluster is bound to an agency in the O&M management area.

- If yes, go to **Step 7**.

- If no, go to **Step 6**.

Step 6 Click **Manage Agency**. On the page that is displayed, rebind the cluster to an agency. Then check whether the alarm is cleared a few minutes later.

- If yes, no further action is required.
- If no, go to **Step 7**.

Step 7 Contact O&M personnel.

----End

9.13.68 ALM-12180 Suspended Disk I/O

Description

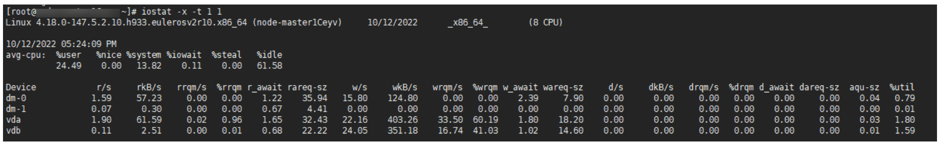
- For HDDs, the alarm is triggered when any of the following conditions is met:
  - The system collects data every 3 seconds, and detects that the **svctm** value exceeds 6s for 10 consecutive periods within 30 seconds.
  - The system collects data every 3 seconds, and detects that the **avgqu-sz** value is greater than 0, the IOPS or bandwidth is 0, and the **ioutil** value is greater than **99%** for 10 consecutive periods within 30 seconds.
- For SSDs, the alarm is triggered when any of the following conditions is met:
  - The system collects data every 3 seconds, and detects that the **svctm** value exceeds 3s for 10 consecutive periods within 30 seconds.
  - The system collects data every 3 seconds, and detects that the **avgqu-sz** value is greater than 0, the IOPS or bandwidth is 0, and the **ioutil** value is greater than **99%** for 10 consecutive periods within 30 seconds.

This alarm is automatically cleared when the preceding conditions have not been met for 90s.

NOTE

- Run the following command in the OS to collect data:

  iostat -x -t 1 1

```
[root@[redacted]~]# iostat -x -t 1 1
Linux 4.18.0-147.5.2.10.h933.eulerosv2r10.x86_64 (node-master1Ceyv)     10/12/2022      _x86_64_        (8 CPU)

10/12/2022 05:24:09 PM
avg-cpu:  %user   %nice %system %iowait  %steal   %idle
          24.49    0.00   13.82    0.11    0.00   61.58

Device            r/s     rkB/s   rrqm/s  %rrqm r_await rareq-sz     w/s     wkB/s   wrqm/s  %wrqm w_await wareq-sz     d/s     dkB/s   drqm/s  %drqm d_await dareq-sz  aqu-sz  %util
dm-0             1.59     57.23     0.00   0.00    1.22    35.94   15.80    124.80     0.00   0.00    2.39     7.90    0.00      0.00     0.00   0.00    0.00     0.00    0.04   0.79
dm-1             0.07      0.30     0.00   0.00    0.67     4.41    0.00      0.00     0.00   0.00    0.00     0.00    0.00      0.00     0.00   0.00    0.00     0.00    0.00   0.01
vda              1.90     61.59     0.02   0.96    1.65    32.43   22.16    403.26    33.50  60.19    1.80    18.20    0.00      0.00     0.00   0.00    0.00     0.00    0.03   1.80
vdb              0.11      2.51     0.00   0.01    0.68    22.22   24.05    351.18    16.74  41.03    1.02    14.60    0.00      0.00     0.00   0.00    0.00     0.00    0.01   1.59
```

  Parameters are as follows:

  avgqu-sz indicates the disk queue depth.

  The sum of **r/s** and **w/s** is the IOPS.

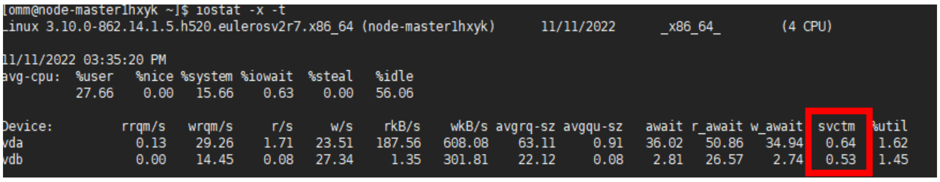
  The sum of **rkB/s** and **wkB/s** is the bandwidth.

  %util is the **ioutil** value.

  The **svctm** value can be obtained as follows:

- MRS 3.1.0:

  Run the **iostat -x -t** command in the OS.

```
[omm@node-master1hxyk ~]$ iostat -x -t
Linux 3.10.0-862.14.1.5.h520.eulerosv2r7.x86_64 (node-master1hxyk)      11/11/2022      _x86_64_        (4 CPU)

11/11/2022 03:35:20 PM
avg-cpu:  %user   %nice %system %iowait  %steal   %idle
          27.66    0.00   15.66    0.63    0.00   56.06

Device:         rrqm/s   wrqm/s     r/s     w/s    rkB/s    wkB/s avgrq-sz avgqu-sz   await r_await w_await  svctm  %util
vda               0.13    29.26    1.71   23.51   187.56   608.08    63.11     0.91   36.02   50.86   34.94   0.64   1.62
vdb               0.00    14.45    0.08   27.34     1.35   301.81    22.12     0.08    2.81   26.57    2.74   0.53   1.45
```

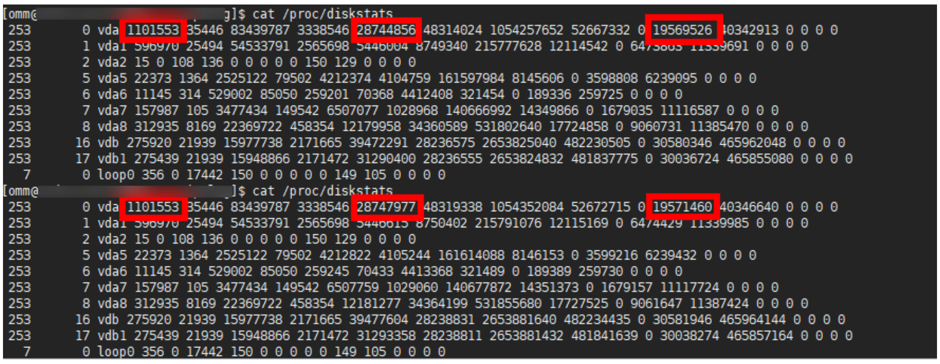
- Calculate **svctm** as follows in versions later than MRS 3.1.0:

  svctm = (tot_ticks_new - tot_ticks_old)/(rd_ios_new + wr_ios_new - rd_ios_old - wr_ios_old)

  If **rd_ios_new + wr_ios_new - rd_ios_old - wr_ios_old** is **0**, then **svctm** is **0**.

  The parameters can be obtained as follows:

  The system runs the **cat /proc/diskstats** command every 3 seconds to collect data. For example:

```
[omm@[redacted]g]$ cat /proc/diskstats
 253       0 vda 1101553 35446 83439787 3338546 28744856 48314024 1054257652 52667332 0 19569526 40342913 0 0 0 0
 253       1 vda1 596970 25494 54533791 2565698 5446004 8749340 215777628 12114542 0 6473803 11339691 0 0 0 0
 253       2 vda2 15 0 108 136 0 0 0 0 0 150 129 0 0 0 0
 253       5 vda5 22373 1364 2525122 79502 4212374 4104759 161597984 8145606 0 3598808 6239095 0 0 0 0
 253       6 vda6 11145 314 529002 85050 259201 70368 4412408 321454 0 189336 259725 0 0 0 0
 253       7 vda7 157987 105 3477434 149542 6507077 1028968 140666992 14349866 0 1679035 11116587 0 0 0 0
 253       8 vda8 312935 8169 22369722 458354 12179958 34360589 531802640 17724858 0 9060731 11385470 0 0 0 0
 253      16 vdb 275920 21939 15977738 2171665 39472291 28236575 2653825040 482230505 0 30580346 465962048 0 0 0 0
 253      17 vdb1 275439 21939 15948866 2171472 31290400 28236555 2653824832 481837775 0 30036724 465855080 0 0 0 0
   7       0 loop0 356 0 17442 150 0 0 0 0 0 149 105 0 0 0 0
[omm@[redacted]]$ cat /proc/diskstats
 253       0 vda 1101553 35446 83439787 3338546 28747977 48319338 1054352084 52672715 0 19571460 40346640 0 0 0 0
 253       1 vda1 596970 25494 54533791 2565698 5446815 8750402 215791076 12115169 0 6474429 11339985 0 0 0 0
 253       2 vda2 15 0 108 136 0 0 0 0 0 150 129 0 0 0 0
 253       5 vda5 22373 1364 2525122 79502 4212822 4105244 161614088 8146153 0 3599216 6239432 0 0 0 0
 253       6 vda6 11145 314 529002 85050 259245 70433 4413368 321489 0 189389 259730 0 0 0 0
 253       7 vda7 157987 105 3477434 149542 6507759 1029060 140677872 14351373 0 1679157 11117724 0 0 0 0
 253       8 vda8 312935 8169 22369722 458354 12181277 34364199 531855680 17727525 0 9061647 11387424 0 0 0 0
 253      16 vdb 275920 21939 15977738 2171665 39477604 28238831 2653881640 482234435 0 30581946 465964144 0 0 0 0
 253      17 vdb1 275439 21939 15948866 2171472 31293358 28238811 2653881432 481841639 0 30038274 465857164 0 0 0 0
   7       0 loop0 356 0 17442 150 0 0 0 0 0 149 105 0 0 0 0
```

  In these two commands:

  In the data collected for the first time, the number in the fourth column is the **rd_ios_old** value, the number in the eighth column is the **wr_ios_old** value, and the number in the thirteenth column is the **tot_ticks_old** value.

  In the data collected for the second time, the number in the fourth column is the **rd_ios_new** value, the number in the eighth column is the **wr_ios_new** value, and the number in the thirteenth column is the **tot_ticks_new** value.

  In this case, the value of **svctm** is as follows:

  (19571460 - 19569526)/(1101553 + 28747977 - 1101553 - 28744856) = 0.6197

Attribute

Alarm ID	Alarm Severity	Auto Clear
12180	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
DiskName	Specifies the disk for which the alarm is generated.

Impact on the System

A continuously high I/O usage may adversely affect service operations and result in service loss.

Possible Causes

The disk is aged.

Procedure

Replace the disk.

Step 1 Log in to FusionInsight Manager and choose **O&M** > **Alarm** > **Alarms**.

Step 2 View the detailed information about the alarm. Check the values of **HostName** and **DiskName** in the location information to obtain the information about the faulty disk for which the alarm is reported.

Step 3 Replace the hard disk.

Step 4 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Collect fault information.

Step 5 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 6 Select **OMS** for **Service** and click **OK**.

Step 7 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 8 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.69 ALM-13000 ZooKeeper Service Unavailable

Description

The system checks the ZooKeeper service status every 60 seconds. This alarm is generated when the ZooKeeper service is unavailable.

This alarm is cleared when the ZooKeeper service recovers.

Attribute

Alarm ID	Alarm Severity	Auto Clear
13000	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

ZooKeeper cannot provide coordination services for upper layer components and the components that depend on ZooKeeper may not run properly.

Possible Causes

- The DNS is installed on the ZooKeeper node.
- The network is faulty.
- The KrbServer service is abnormal.
- The ZooKeeper instance is abnormal.
- The disk capacity is insufficient.

Procedure

Check the DNS.

Step 1 Check whether the DNS is installed on the node where the ZooKeeper instance is located. On the Linux node where the ZooKeeper instance is located, run the **cat /etc/resolv.conf** command to check whether the file is empty.

- If yes, go to **Step 2**.
- If no, go to **Step 3**.

Step 2 Run the **service named status** command to check whether the DNS is started.

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 Run the **service named stop** command to stop the DNS service. If "Shutting down name server BIND waiting for named to shut down (28s)" is displayed, the DNS service is stopped successfully. Comment out the content (if any) in **/etc/resolv.conf**.

Step 4 On the **O&M > Alarm > Alarms** tab, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check the network status.

Step 5 On the Linux node where the ZooKeeper instance is located, run the **ping** command to check whether the host names of other nodes where the ZooKeeper instance is located can be pinged successfully.

- If yes, go to **Step 9**.
- If no, go to **Step 6**.

Step 6 Modify the IP addresses in **/etc/hosts** and add the host name and IP address mapping.

Step 7 Run the **ping** command again to check whether the host names of other nodes where the ZooKeeper instance is located can be pinged successfully.

- If yes, go to **Step 8**.
- If no, go to **Step 23**.

Step 8 On the **O&M** > **Alarm** > **Alarms** tab, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Check the KrbServer service status (Skip this step if the normal mode is used).

Step 9 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 10 Check whether the KrbServer service is normal.

- If yes, go to **Step 13**.
- If no, go to **Step 11**.

Step 11 Perform operations based on "ALM-25500 KrbServer Service Unavailable" and check whether the KrbServer service is recovered.

- If yes, go to **Step 12**.
- If no, go to **Step 23**.

Step 12 On the **O&M** > **Alarm** > **Alarms** tab, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 13**.

Check the ZooKeeper service instance status.

Step 13 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **ZooKeeper** > **quorumpeer**.

Step 14 Check whether the ZooKeeper instances are normal.

- If yes, go to **Step 18**.
- If no, go to **Step 15**.

Step 15 Select instances whose status is not good, and choose **More** > **Restart Instance**.

Step 16 Check whether the instance status is good after restart.

- If yes, go to **Step 17**.
- If no, go to **Step 18**.

Step 17 On the **O&M** > **Alarm** > **Alarms** tab, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 18**.

Check disk status.

Step 18 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Service** > **ZooKeeper** > **quorumpeer**, and check the node host information of the ZooKeeper instance.

Step 19 On FusionInsight Manager, click **Host**.

Step 20 In the **Disk** column, check whether the disk space of each node where ZooKeeper instances are located is insufficient (disk usage exceeds 80%).

- If yes, go to **Step 21**.

- If no, go to **Step 23**.

Step 21 Expand disk capacity. For details, see "ALM-12017 Insufficient Disk Capacity".

Step 22 On the **O&M** > **Alarm** > **Alarms** tab, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 23**.

Collect fault information.

Step 23 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 24 Select the following nodes in the required cluster from the **Service**: (KrbServer logs do not need to be downloaded in normal mode.)

- ZooKeeper
- KrbServer

Step 25 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 26 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.70 ALM-13001 Available ZooKeeper Connections Are Insufficient

Description

The system checks ZooKeeper connections every 60 seconds. This alarm is generated when the system detects that the number of used ZooKeeper instance connections exceeds the threshold (80% of the maximum connections).

When the **Trigger Count** is 1, this alarm is cleared when the number of used ZooKeeper instance connections is smaller than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the number of used ZooKeeper instance connections is smaller than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
13001	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host name for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

Available ZooKeeper connections are insufficient. When the connection usage reaches 100%, external connections cannot be handled.

Possible Causes

The number of connections to the ZooKeeper node exceeds the threshold. Connection leakage occurs on some connection processes, or the maximum number of connections does not comply with the actual scenario.

Procedure

Check connection status.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms**. On the displayed interface, click the drop-down button of **Available ZooKeeper Connections Are Insufficient** and confirm the node IP address of the host for which the alarm is generated in the Location Information.

Step 2 Obtain the PID of the ZooKeeper process. Log in to the node involved in this alarm as user **root** and run the **pgrep -f proc_zookeeper** command.

Step 3 Check whether the PID can be correctly obtained.

- If yes, go to **Step 4**.
- If no, go to **Step 15**.

Step 4 Obtain all the IP addresses connected to the ZooKeeper instance and the number of connections and check 10 IP addresses with top connections. Run the following command based on the obtained PID: **lsof -i|grep** *$pid* **| awk '{print $9}' | cut -d :**

-f 2 | cut -d \>-f 2 | awk '{a[$1]++} END {for(i in a){print i,a[i] | "sort -r -g -k 2"}}' | head -10. (The PID obtained in the preceding step is used.)

Step 5 Check whether node IP addresses and number of connections are successfully obtained.

- If yes, go to **Step 6**.
- If no, go to **Step 15**.

Step 6 Obtain the ID of the port connected to the process. Run the following command based on the obtained PID and IP address: **lsof -i|grep** *$pid* **| awk '{print $9}'|cut -d \> -f 2 |grep** *$IP***| cut -d : -f 2**. (The PID and IP address obtained in the preceding step are used.)

Step 7 Check whether the port ID is successfully obtained.

- If yes, go to **Step 8**.
- If no, go to **Step 15**.

Step 8 Obtain the ID of the connected process. Log in to each IP address and run the following command based on the obtained port ID: **lsof -i|grep** *$port*. (The port ID obtained in the preceding step is used.)

Step 9 Check whether the process ID is successfully obtained.

- If yes, go to **Step 10**.
- If no, go to **Step 15**.

Step 10 Check whether connection leakage occurs on the process based on the obtained process ID.

- If yes, go to **Step 11**.
- If no, go to **Step 12**.

Step 11 Close the process where connection leakage occurs and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 12**.

Step 12 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **ZooKeeper** > **Configurations** > **All Configurations** > **quorumpeer** > **Performance** and increase the value of **maxCnxns** as required.

Step 13 Save the configuration and restart the ZooKeeper service.

Step 14 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 15**.

Collect fault information.

Step 15 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 16 Select **ZooKeeper** in the required cluster from the **Service**:

Step 17 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 18 Contact the O&M personnel and send the collected log information.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.71 ALM-13002 ZooKeeper Direct Memory Usage Exceeds the Threshold

Description

The system checks the direct memory usage of the ZooKeeper service every 30 seconds. The alarm is generated when the direct memory usage of a ZooKeeper instance exceeds the threshold (80% of the maximum memory).

When the **Trigger Count** is 1, this alarm is cleared when the ZooKeeper Direct memory usage is less than the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the ZooKeeper Direct memory usage is less than 80% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
13002	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available direct memory of the ZooKeeper service is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The direct memory of the ZooKeeper instance is overused or the direct memory is inappropriately allocated.

Procedure

Check the direct memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms**. On the displayed interface, click the drop-down button of **ZooKeeper Direct Memory Usage Exceeds the Threshold**. Check the IP address of the instance that reports the alarm.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **ZooKeeper** > **Instance** > **quorumpeer(the IP address checked)**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **CPU and Memory**, and select **ZooKeeper Heap And Direct Buffer Resource Percentage**, click **OK**.

Step 3 Check whether the used direct buffer memory of ZooKeeper reaches 80% of the maximum direct buffer memory specified for ZooKeeper.

- If yes, go to **Step 4**.
- If no, go to **Step 8**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **ZooKeeper** > **Configurations** > **All Configurations** > **quorumpeer** > **System** to check whether "-XX:MaxDirectMemorySize" exists in the **GC_OPTS** parameter.

- If yes, in the **GC_OPTS** parameter, delete "-XX:MaxDirectMemorySize" and go to **Step 5**.
- If no, go to **Step 6**.

Step 5 Save the configuration and restart the ZooKeeper service.

Step 6 Check whether the **ALM-13004 ZooKeeper Heap Memory Usage Exceeds the Threshold** exists.

- If yes, handle the alarm by referring to **ALM-13004 ZooKeeper Heap Memory Usage Exceeds the Threshold**.
- If no, go to **Step 7**.

Step 7 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Collect fault information.

Step 8 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 9 Select **ZooKeeper** in the required cluster from the **Service**.

Step 10 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.72 ALM-13003 GC Duration of the ZooKeeper Process Exceeds the Threshold

Description

The system checks the garbage collection (GC) duration of the ZooKeeper process every 60 seconds. This alarm is generated when the GC duration exceeds the threshold (12 seconds by default).

This alarm is cleared when the GC duration is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
13003	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Name	Meaning
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

A long GC duration of the ZooKeeper process may interrupt the services.

Possible Causes

The heap memory of the ZooKeeper process is overused or inappropriately allocated, causing frequent occurrence of the GC process.

Procedure

Check the GC duration.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. On the displayed page, click the drop-down list of **GC Duration of the ZooKeeper Process Exceeds the Threshold**. View the IP address of the instance for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **ZooKeeper** > **Instance** > **quorumpeer**. Click the drop-down list in the upper right corner of **Chart**, choose **Customize** > **GC**, select **ZooKeeper GC Duration per Minute**, and click **OK** to check the GC duration statistics of the ZooKeeper process collected every minute.

Step 3 Check whether the GC duration of the ZooKeeper process collected every minute exceeds the threshold (12 seconds by default).

- If yes, go to **Step 4**.
- If no, go to **Step 8**.

Step 4 Check whether memory leakage occurs in the application.

Step 5 On the **Home** page of FusionInsight Manager, choose **Cluster** > **Services** > **ZooKeeper**. On the page that is displayed, click the **Configuration** tab then the **All Configurations** sub-tab, and select **quorumpeer** > **System**. Increase the value of the **GC_OPTS** parameter as required.

NOTE

Generally, **-Xmx** is twice of ZooKeeper data capacity. If the capacity of ZooKeeper reaches 2 GB, set **GC_OPTS** as follows:

-Xms4G -Xmx4G -XX:NewSize=512M -XX:MaxNewSize=512M -XX:MetaspaceSize=64M -XX:MaxMetaspaceSize=64M -XX:CMSFullGCsBeforeCompaction=1

Step 6 Save the configuration and restart the ZooKeeper service.

Step 7 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Collect the fault information.

Step 8 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 9 Expand the **Service** drop-down list, and select **ZooKeeper** for the target cluster.

Step 10 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.73 ALM-13004 ZooKeeper Heap Memory Usage Exceeds the Threshold

Description

The system checks the heap memory usage of the ZooKeeper service every 60 seconds. The alarm is generated when the heap memory usage of a ZooKeeper instance exceeds the threshold (95% of the maximum memory).

The alarm is cleared when the memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
13004	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.

Name	Meaning
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available ZooKeeper heap memory is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The heap memory of the ZooKeeper instance is overused or the heap memory is inappropriately allocated.

Procedure

Check heap memory usage.

Step 1 On the FusionInsight Manager portal, On the displayed interface, click the drop-down button of **ZooKeeper Heap Memory Usage Exceeds the Threshold** and confirm the node IP address of the host for which the alarm is generated in the Location Information.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **ZooKeeper** > **Instance**, click **quorumpeer** in the **Role** column of the corresponding IP address. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **CPU and Memory**, and select **ZooKeeper Heap And Direct Buffer Resource Percentage**, click **OK**. Check the heap memory usage.

Step 3 Check whether the used heap memory of ZooKeeper reaches 95% of the maximum heap memory specified for ZooKeeper.

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **ZooKeeper** > **Configurations** > **All Configurations** > **quorumpeer** > **System**. Increase the value of **-Xmx** in **GC_OPTS** as required. The details are as follows:

1. On the **Instance** tab, click **quorumpeer** in the **Role** column of the corresponding IP address. Choose **Customize** > **CPU and Memory** in the upper right corner, and select **ZooKeeper Heap And Direct Buffer Resource**, click **OK** to check the heap memory used by ZooKeeper.
2. Change the value of **-Xmx** in the **GC_OPTS** parameter based on the actual heap memory usage. Generally, the value is twice the size of the ZooKeeper

data volume. For example, if 2 GB ZooKeeper heap memory is used, the following configurations are recommended: -Xms4G -Xmx4G -XX:NewSize=512M -XX:MaxNewSize=512M -XX:MetaspaceSize=64M -XX:MaxMetaspaceSize=64M -XX:CMSFullGCsBeforeCompaction=1

Step 5 Save the configuration and restart the ZooKeeper service.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **ZooKeeper** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.74 ALM-13005 Failed to Set the Quota of Top Directories of ZooKeeper Components

Description

The system sets quotas for each ZooKeeper top-level directory in the **customized.quota** configuration item and components every 5 hours. This alarm is generated when the system fails to set the quota for a directory.

This alarm is cleared when the setting succeeds after a failure.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
13005	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
ServiceDirectory	Specifies the directory for which the alarm is generated.
Trigger Condition	Specifies the cause of the alarm.

Impact on the System

Components can write a large amount of data to the top-level directory of ZooKeeper. As a result, the ZooKeeper service is unavailable.

Possible Causes

The quota for the alarm directory is inappropriate.

Procedure

Check whether the quota for the alarm directory is appropriate.

Step 1 Log in to FusionInsight Manager, and choose **Cluster >** *Name of the desired cluster* **> Services** > **ZooKeeper**. On the displayed page, choose **Configurations** > **All Configurations** > **Quota**. Check whether the directory for which the alarm is reported and its quota exist in the **customized.quota** configuration item.

- If yes, go to **Step 5**.
- If no, go to **Step 2**.

Step 2 Check whether the alarm directory for which the alarm is reported is in the following alarm list.

Table 9-89 Component alarm directory

Component	Alarm Directory
Hbase	/hbase
Hive	/beelinesql
Yarn	/rmstore
Storm	/stormroot
Streaming	/storm
Kafka	/kafka

- If yes, go to **Step 3**.
- If no, go to **Step 7**.

Step 3 View the component of the alarm directory in the table, open the corresponding service page, and choose **Configurations** > **All Configurations**. On the displayed page, search for **zk.quota** in the upper right corner. The search result is the quota of the alarm directory.

Step 4 Check whether the quota of the alarm directory for which the alarm is reported is appropriate. The quota must be greater than or equal to the actual value, which can be obtained in **Trigger Condition**.

Step 5 Modify the **services.quota** value as prompted and save the configuration.

Step 6 After the time specified by **service.quotas.auto.check.cron.expression**, check whether the alarm is cleared.

- If it is, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **ZooKeeper** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.75 ALM-13006 Znode Number or Capacity Exceeds the Threshold

Description

The system periodically detects the status of secondary Znode in the ZooKeeper service data directory every four hours. This alarm is generated when the number or capacity of secondary Znodes exceeds the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
13006	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
ServiceDirectory	Specifies the directory for which the alarm is generated.
Trigger Condition	Specifies the cause of the alarm.

Impact on the System

A large amount of data is written to the ZooKeeper data directory. As a result, ZooKeeper cannot provide normal services.

Possible Causes

A large amount of data is written to the ZooKeeper data directory. The threshold is not appropriate.

Procedure

Check whether a large amount of data is written to the directory for which the alarm is generated.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. On the displayed interface, click the drop-down button of **Znode Number or Capacity Exceeds the Threshold**. Confirm the Znode for which the alarm is generated in Location Information.

Step 2 Log in to FusionInsight Manager, open the ZooKeeper service interface, and select **Resource**. In the table **Used Resources (By Second-Level Znode)**, check whether a large amount of data is written to the top-level Znode for which the alarm is reported.

- If it is, go to **Step 3**.
- If it is not, go to **Step 4**.

Step 3 Log in to the ZooKeeper client and delete the data in the top-level Znode.

Step 4 Log in to FusionInsight Manager and open the ZooKeeper service interface. On the **Resource** page, choose ◷ > **By Znode quantity** in **Used Resources (By Second-**

Level Znode). **Threshold Configuration of By Znode quantity** is displayed. Click **Modify** under **Operation**. Increase the threshold by referringto the value of **max.Znode.count** by choosing **Cluster >** *Name of the desired cluster* **> Services** > **ZooKeeper** > **Configurations > All Configurations > Quota**.

Step 5 In the **Used Resources (By Second-Level Znode)**, choose > **By capacity**. The **Threshold Settings** page of **By Capacity** is displayed. Click **Modify** under **Operation**. Increase the threshold by referring to the value of **max.data.size** by choosing **Cluster >** *Name of the desired cluster* **> Services** > **ZooKeeper** > **Configurations > All Configurations** > **Quota**.

Step 6 Check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **ZooKeeper** in the required cluster from the **Service**.

Step 9 Click in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.76 ALM-13007 Available ZooKeeper Client Connections Are Insufficient

Description

The system periodically detects the number of active processes between the ZooKeeper client and the ZooKeeper server every 60 seconds. This alarm is generated when the number of connections exceeds the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
13007	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the host name for which the alarm is generated.
ClientIP	Specifies the client IP address.
ServerIP	Specifies the server IP address.
Trigger Condition	Specifies the cause of the alarm.

Impact on the System

A large number of connections to ZooKeeper caused the ZooKeeper to be fully connected and unable to provide normal services.

Possible Causes

A large number of client processes are connected to ZooKeeper. The thresholds are not appropriate.

Procedure

Check whether there are a large number of client processes connected to ZooKeeper.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. On the displayed interface, click the drop-down button of **Available ZooKeeper Client Connections Are Insufficient**. Confirm the node IP address of the host for which the alarm is generated in the Location Information.

Step 2 Open the ZooKeeper service interface, click **Resource** to enter the **Resource** page, and check whether the number of connections of the client with the IP address specified by **Number of Connections (By Client IP Address)** is large.

- If it is, go to **Step 3**.
- If it is not, go to **Step 4**.

Step 3 Check whether connection leakage occurs on the client process.

Step 4 Click in the **Number of Connections (by Client IP Address)** to enter the **Thresholds** page, and click **Modify** under **Operation**. Increase the threshold by referring to the value of **maxClientCnxns** by choosing **Cluster >** *Name of the* *desired cluster* **> Services** > **ZooKeeper** > **Configurations > All Configurations > quorumpeer**.

Step 5 Check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 7 Select **ZooKeeper** in the required cluster from the **Service**.

Step 8 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.77 ALM-13008 ZooKeeper Znode Usage Exceeds the Threshold

Description

The system checks the level-2 Znode status in the ZooKeeper data directory every hour. This alarm is generated when the system detects that the level-2 Znode usage exceeds the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
13008	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service name for which the alarm is generated.
ServiceDirectory	Specifies the directory for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
Trigger Condition	Specifies the cause of the alarm.

Impact on the System

A large amount of data is written to the ZooKeeper data directory. As a result, ZooKeeper cannot provide services properly.

Possible Causes

- A large amount of data is written to the ZooKeeper data directory.
- The user-defined threshold is inappropriate.

Procedure

Check whether a large amount of data is written into the directory for which the alarm is generated.

Step 1 Log in to FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **ZooKeeper**, and click **Resource**. Click **By Znode quantity** in **Used Resources (By Second-Level Znode)**, and check whether a large amount of data is written to the top Znode.

- If yes, go to **Step 2**.
- If no, go to **Step 4**.

Step 2 Log in to FusionInsight Manager, choose **O&M > Alarm > Alarms**, select **Location** from the drop-down list box next to **ALM-13008 ZooKeeper Znode Quantity Usage Exceeds Threshold**, and obtain the Znode path in **ServiceDirectory**.

Step 3 Log in to the ZooKeeper client as a cluster user and delete unnecessary data from the Znode corresponding to the alarm.

Step 4 Log in to FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **ZooKeeper** > **Configurations** > **All Configurations**, and search for **max.znode.count**, which is the maximum number of ZooKeeper directories. The alarm threshold is 80% of this parameter. Increase the value of this parameter, click **Save**, and restart the service for the configuration to take effect.

Step 5 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 7 Select **ZooKeeper** in the required cluster from the **Service**.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.78 ALM-13009 ZooKeeper Znode Capacity Usage Exceeds the Threshold

Description

The system checks the level-2 ZNode status in the ZooKeeper data directory every hour. This alarm is generated when the system detects that the capacity usage exceeds the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
13009	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
ServiceDirectory	Specifies the directory for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.

Name	Meaning
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

A large amount of data is written to the ZooKeeper data directory. As a result, ZooKeeper cannot provide services properly.

Possible Causes

- A large volume of data has been written to the ZooKeeper data directory.
- The threshold is improperly defined.

Procedure

Check whether a large volume of data is written to the alarm directory.

Step 1 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Alarm** > **Alarms**. Click the drop-down list in the row containing **ALM-13009 ZooKeeper ZNode Capacity Usage Exceeds the Threshold**, and find the ZNode for which the alarm is generated in the **Location** area.

Step 2 Choose **Cluster** > **Services** > **ZooKeeper**. On the page that is displayed, click the **Resource** tab. In the **Used Resources (By Second-Level ZNode)** area, click **By capacity** and check whether a large amount of data is written to the top-level ZNode directory.

- If yes, record the directory to which a large amount of data is written and go to **Step 3**.
- If no, go to **Step 5**.

Step 3 Check whether data in the directory can be deleted.

NOTICE

Deleting data from ZooKeeper is a high-risk operation. Exercise caution when performing this operation.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 Log in to the ZooKeeper client and delete unnecessary data from the directory to which a large amount of data is written.

1. Log in to the ZooKeeper client installation directory, for example, **/opt/client**, and configure environment variables.

   ```
   cd /opt/client
   source bigdata_env
   ```

2. Run the following command to authenticate the user (skip this step for a cluster in normal mode):

   kinit *Component service user*

3. Run the following command to log in to the client tool:

   zkCli.sh -server <*Service IP address of the node where any ZooKeeper instance resides***>:<***Client port***>**

4. Run the following command to delete unnecessary data:

   delete *Path of the file to be deleted*

Step 5 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **ZooKeeper**. On the page that is displayed, click the **Configuration** tab then the **All Configurations** sub-tab, and search for **max.data.size**. The value of **max.data.size** is the maximum capacity quota of the ZooKeeper directory. The unit is byte. Search for the **GC_OPTS** configuration item and check the value of **Xmx**.

Step 6 Compare the values of **max.data.size** and **Xmx*0.65**. The threshold is the smaller value multiplied by 80%. You can change the values of **max.data.size** and **Xmx*0.65** to increase the threshold.

Step 7 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Collect the fault information.

Step 8 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 9 Expand the **Service** drop-down list, and select **ZooKeeper** for the target cluster.

Step 10 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.79 ALM-13010 Znode Usage of a Directory with Quota Configured Exceeds the Threshold

Description

The system checks the Znode usage of all service directories with quota configured every hour. This alarm is generated when the system detects that the level-2 Znode usage exceeds the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
13010	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
ServiceDirectory	Specifies the directory for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
Trigger Condition	Specifies the cause of the alarm.

Impact on the System

A large amount of data is written to the ZooKeeper data directory. As a result, ZooKeeper cannot provide services properly.

Possible Causes

- A large amount of data is written to the ZooKeeper data directory.
- The user-defined threshold is inappropriate.

Procedure

Check whether a large amount of data is written into the directory for which the alarm is generated.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm > Alarms**. Confirm the Znode for which the alarm is generated in **Location** of this alarm.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services** > **ZooKeeper** and click **Resource**. In **Used Resources (By Second-Level Znode)**, check whether a large amount of data is written into the top Znode.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 3 Log in to FusionInsight Manager, choose **O&M > Alarm > Alarms**, select Location from the drop-down list box next to **ALM-13010 Znode Usage of a Directory with Quota Configured Exceeds the Threshold**, and obtain the Znode path in ServiceDirectory.

Step 4 Log in to the ZooKeeper client as a cluster user and delete unwanted data in the Znode for which the alarm is generated.

Step 5 Log in to FusionInsight Manager, and choose **Cluster** > *Name of the desired cluster* > **Services** > *Component of the top Znode for which the alarm isgenerated.* Choose **Configurations** > **All Configurations**, search for **zk.quota.number**, increase its value, click **Save**.

NOTICE

If the Component of the top Znode for which the alarm isgenerated is ClickHouse, change the value of **clickhouse.zookeeper.quota.node.count**.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **ZooKeeper** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.80 ALM-14000 HDFS Service Unavailable

Description

The system checks the NameService service status every 60 seconds. This alarm is generated when all the NameService services are abnormal and the system considers that the HDFS service is unavailable.

This alarm is cleared when at least one NameService service is normal and the system considers that the HDFS service recovers.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14000	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

HDFS fails to provide services for HDFS service-based upper-layer components, such as HBase and MapReduce. As a result, users cannot read or write files.

Possible Causes

- The ZooKeeper service is abnormal.
- All NameService services are abnormal.

Procedure

Check the ZooKeeper service status.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms**. On the Alarm page, check whether **ALM-13000 ZooKeeper Service Unavailable** is reported.

- If yes, go to **Step 2**.
- If no, go to **Step 4**.

Step 2 See **ALM-13000 ZooKeeper Service Unavailable** to rectify the health status of ZooKeeper fault and check whether the **Running Status** of the ZooKeeper service restores to **Normal**.

- If yes, go to **Step 3**.
- If no, go to **Step 7**.

Step 3 On the **O&M** > **Alarm** > **Alarms** page, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Handle the NameService service exception alarm.

Step 4 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms**. On the Alarms page, check whether **ALM-14010 NameService Service Unavailable** is reported.

- If yes, go to **Step 5**.
- If no, go to **Step 7**.

Step 5 See **ALM-14010 NameService Service Unavailable** to handle the abnormal NameService services and check whether each NameService service exception alarm is cleared.

- If yes, go to **Step 6**.
- If no, go to **Step 7**.

Step 6 On the **O&M** > **Alarm** > **Alarms** page, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 8 Select the following nodes in the required cluster from the **Service**:

- ZooKeeper
- HDFS

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.81 ALM-14001 HDFS Disk Usage Exceeds the Threshold

Description

The system checks the HDFS disk usage every 30 seconds and compares the actual HDFS disk usage with the threshold. The HDFS disk usage indicator has a default threshold, this alarm is generated when the value of the disk usage of a Hadoop distributed file system (HDFS) indicator exceeds the threshold.

To change the threshold, choose **O&M** > **Alarm > Thresholds** > *Name of the desired cluster* **> HDFS**.

When the **Trigger Count** is 1, this alarm is cleared when the value of the disk usage of HDFS cluster indicator is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the value of the disk usage of HDFS cluster indicator is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14001	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
NameServiceName	Specifies the NameService for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

Writing Hadoop distributed file system (HDFS) data is affected.

Possible Causes

The disk space configured for the HDFS cluster is insufficient.

Procedure

Check the disk capacity and delete unnecessary files.

Step 1 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS**.

Step 2 Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **Disk**, and select **Percentage of HDFS Capacity** to check whether the HDFS disk usage exceeds the threshold (80% by default).

- If yes, go to **Step 3**.
- If no, go to **Step 11**.

Step 3 In the **Basic Information** area, click the **NameNode(Active)** of the failure NameService and the HDFS WebUI page is displayed.

NOTE

By default, the **admin** user does not have the permissions to manage other components. If the page cannot be opened or the displayed content is incomplete when you access the native UI of a component due to insufficient permissions, you can manually create a user with the permissions to manage that component.

Step 4 On the HDFS web user interface (WebUI), click **Datanodes** tab. In the **Block pool used** column, view the disk usage of all DataNodes to check whether the disk usage of any DataNode exceeds the threshold.

- If yes, go to **Step 6**.
- If no, go to **Step 11**.

Step 5 Log in to the MRS client node as user **root**.

Step 6 Run **cd /opt/client** to switch to the client installation directory, and run **source bigdata_env**. If the cluster uses the security mode, perform security authentication. Run **kinit hdfs** and enter the password as prompted. Please obtain the password from the administrator.

Step 7 Run the **hdfs dfs -rm -r** *file or directory* command to delete unnecessary files.

Step 8 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Expand the system.

Step 9 Expand the disk capacity.

Step 10 Check whether the alarm is cleared.

- If yes, no further action is required.

- If no, go to **Step 11**.

Collect fault information.

Step 11 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 12 Select the following nodes in the required cluster from the **Service**:

- ZooKeeper
- HDFS

Step 13 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 14 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.82 ALM-14002 DataNode Disk Usage Exceeds the Threshold

Description

The system checks the DataNode disk usage every 30 seconds and compares the actual disk usage with the threshold. A default threshold range is provided for the DataNode disk usage. This alarm is generated when the DataNode disk usage exceeds the threshold.

To change the threshold, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **HDFS**.

If **Trigger Count** is **1**, this alarm is cleared when the DataNode disk usage is less than or equal to the threshold. If **Trigger Count** is greater than **1**, this alarm is cleared when the DataNode disk usage is less than or equal to 80% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
14002	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Insufficient disk space will impact data write to HDFS.

Possible Causes

- The disk space configured for the HDFS cluster is insufficient.
- Data skew occurs among DataNodes.

Procedure

Check whether the cluster disk capacity is full.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**, and check whether the **ALM-14001 HDFS Disk Usage Exceeds the Threshold** alarm exists.

- If yes, go to **Step 2**.
- If no, go to **Step 4**.

Step 2 Handle the alarm by following the instructions in **ALM-14001 HDFS Disk Usage Exceeds the Threshold** and check whether the alarm is cleared.

- If yes, go to **Step 3**.
- If no, go to **Step 11**.

Step 3 Choose **O&M** > **Alarm** > **Alarms** and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check the balance status of DataNodes.

Step 4 On FusionInsight Manager, choose **Hosts**. Check whether the number of DataNodes on each rack is almost the same. If the difference is large, adjust the racks to which DataNodes belong to ensure that the number of DataNodes on each rack is almost the same. Restart the HDFS service for the settings to take effect.

Step 5 Choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS**.

Step 6 In the **Basic Information** area, click **NameNode(Active)**. The HDFS web UI is displayed.

NOTE

By default, the **admin** user does not have the permissions to manage other components. If the page cannot be opened or the displayed content is incomplete when you access the native UI of a component due to insufficient permissions, you can manually create a user with the permissions to manage that component.

Step 7 In the **Summary** area of the HDFS web UI, check whether the value of **Max** is 10% greater than that of **Median** in **DataNodes usages**.

- If yes, go to **Step 8**.
- If no, go to **Step 11**.

Step 8 Balance skewed data in the cluster. Log in to the MRS client as user **root**. If the cluster is in normal mode, run the **su - omm** command to switch to user **omm**. Run the **cd** command to go to the client installation directory and run the **source bigdata_env** command. If the cluster uses the security mode, perform security authentication. Run **kinit hdfs** and enter the password as prompted. Obtain the password from the MRS cluster administrator.

Step 9 Run the following command to balance data distribution:

hdfs balancer -threshold 10

Step 10 Wait several minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Collect the fault information.

Step 11 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 12 Expand the drop-down list next to the **Service** field. In the **Services** dialog box that is displayed, select **HDFS** for the target cluster.

Step 13 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 14 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.83 ALM-14003 Number of Lost HDFS Blocks Exceeds the Threshold

Description

The system checks the lost blocks every 30 seconds and compares the actual lost blocks with the threshold. The lost blocks indicator has a default threshold. This alarm is generated when the number of lost HDFS blocks exceeds the threshold.

To change the threshold, choose **O&M** > **Alarm > Thresholds** > *Name of the desired cluster* **> HDFS**.

If **Trigger Count** is **1**, this alarm is cleared when the value of lost HDFS blocks is less than or equal to the threshold. If **Trigger Count** is greater than **1**, this alarm is cleared when the value of lost HDFS blocks is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
14003	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
NameServiceName	Specifies the NameService for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Data stored in HDFS is lost. HDFS may enter the safe mode and cannot provide write services. Lost block data cannot be restored.

Possible Causes

- The DataNode instance is abnormal.
- Data is deleted.

Procedure

Check the DataNode instance.

Step 1 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS** > **Instance**.

Step 2 Check whether the **Running Status** of all DataNode instance is **Normal**.

- If yes, go to **Step 11**.
- If no, go to **Step 3**.

Step 3 Restart the DataNode instance and check whether the DataNode instance restarts successfully.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 Choose **O&M** > **Alarm** > **Alarms** and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Delete the damaged file.

Step 5 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS** > **NameNode(Active)**. On the WebUI page of the HDFS, view the information about lost blocks.

NOTE

- If a block is lost, a line in red is displayed on the WebUI.
- By default, the **admin** user does not have the permissions to manage other components. If the page cannot be opened or the displayed content is incomplete when you access the native UI of a component due to insufficient permissions, you can manually create a user with the permissions to manage that component.

Step 6 The user checks whether the file containing the lost data block is useful.

NOTE

Files generated in directories **/mr-history**, **/tmp/hadoop-yarn**, and **/tmp/logs** during MapReduce task execution are unnecessary.

- If yes, go to **Step 7**.
- If no, go to **Step 8**.

Step 7 The user checks whether the file containing the lost data block is backed up.

- If yes, go to **Step 8**.
- If no, go to **Step 11**.

Step 8 Log in to the HDFS client as user **root**. The user password is defined by the user before the installation. Contact the MRS cluster administrator to obtain the password. Run the following commands:

- Security mode:

  cd *Client installation directory*

  source bigdata_env

  kinit hdfs
- Normal mode:

  su - omm

  cd *Client installation directory*

  source bigdata_env

Step 9 On the node client, run **hdfs fsck / -delete** to delete the lost file. If the file where the lost block is located is a useful file, you need to write the file again to restore the data.

NOTE

Deleting a file or folder is a high-risk operation. Ensure that the file or folder is no longer required before performing this operation.

Step 10 Choose **O&M** > **Alarm** > **Alarms** and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Collect the fault information.

Step 11 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 12 Expand the drop-down list next to the **Service** field. In the **Services** dialog box that is displayed, select **HDFS** for the target cluster.

Step 13 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 14 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.84 ALM-14006 Number of HDFS Files Exceeds the Threshold

Description

The system periodically checks the number of HDFS files every 30 seconds and compares the number of HDFS files with the threshold. This alarm is generated when the system detects that the number of HDFS files exceeds the threshold.

If **Trigger Count** is **1**, this alarm is cleared when the number of HDFS files is less than or equal to the threshold. If **Trigger Count** is greater than **1**, this alarm is cleared when the number of HDFS files is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
14006	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
NameServiceName	Specifies the NameService for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Disk storage space is insufficient, which may result in data import failure. The performance of the HDFS system is affected.

Possible Causes

The number of HDFS files exceeds the threshold.

Procedure

Check the number of files in the system.

Step 1 On FusionInsight Manager, check the number of HDFS files. Specifically, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **File and Block**, and select **HDFS File** and **Total Blocks**.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS** > **Configurations** > **All Configurations**, and search for the **GC_OPTS** parameter under **NameNode**.

Step 3 Configure the threshold of the number of configuration file objects. Specifically, change the value of **Xmx** (GB) in the **GC_OPTS** parameter. The threshold (specified by y) is calculated as follows: y = 0.2007 x Xmx - 0.6312, where x indicates the memory capacity Xmx (GB) and y indicates the number of files (unit: kW). Adjust the memory size as required.

Step 4 Confirm that the value of **GC_PROFILE** is **custom** so that the **GC_OPTS** configuration takes effect. Click **Save** and choose **More** > **Restart Instance** to restart the service.

Step 5 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Check whether needless files exist in the system.

Step 6 Log in to the HDFS client as user **root**. Run **cd** to switch to the client installation directory, and run **source bigdata_env** to configure the environment variables.

If the cluster uses the security mode, perform security authentication.

Run the **kinit hdfs** command and enter the password as prompted. Obtain the password from the MRS cluster administrator.

Step 7 Run **hdfs dfs -ls** *file or directory* to check whether the files in the directory can be deleted.

- If yes, go to **Step 8**.
- If no, go to **Step 9**.

Step 8 Run the **hdfs dfs -rm -r** *file or directory path* command. After deleting unnecessary files, wait until the files are retained in the recycle bin for a period longer than the value of **fs.trash.interval** on the NameNode. Then check whether the alarm is cleared.

NOTE

Deleting a file or folder is a high-risk operation. Ensure that the file or folder is no longer required before performing this operation.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect the fault information.

Step 9 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 10 Expand the drop-down list next to the **Service** field. In the **Services** dialog box that is displayed, select **HDFS** for the target cluster.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

Configuration rules of the NameNode JVM parameter

Default value of the NameNode JVM parameter **GC_OPTS**:

-Xms2G -Xmx4G -XX:NewSize=128M -XX:MaxNewSize=256M -XX:MetaspaceSize=128M -XX:MaxMetaspaceSize=128M -XX:+UseConcMarkSweepGC -XX:+CMSParallelRemarkEnabled -XX:CMSInitiatingOccupancyFraction=65 -XX:+PrintGCDetails -Dsun.rmi.dgc.client.gcInterval=0x7FFFFFFFFFFFFFE -Dsun.rmi.dgc.server.gcInterval=0x7FFFFFFFFFFFFFE -XX:-OmitStackTraceInFastThrow -XX:+PrintGCDateStamps -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=1M -Djdk.tls.ephemeralDHKeySize=3072 -Djdk.tls.rejectClientInitiatedRenegotiation=true -Djava.io.tmpdir=${Bigdata_tmp_dir}

The number of NameNode files is proportional to the used memory size of the NameNode. When file objects change, you need to change **-Xms2G -Xmx4G -XX:NewSize=128M -XX:MaxNewSize=256M** in the default value. The following table lists the reference values.

Table 9-90 NameNode JVM configuration

Number of File Objects	Reference Value
10,000,000	-Xms6G -Xmx6G -XX:NewSize=512M -XX:MaxNewSize=512M
20,000,000	-Xms12G -Xmx12G -XX:NewSize=1G -XX:MaxNewSize=1G
50,000,000	-Xms32G -Xmx32G -XX:NewSize=3G -XX:MaxNewSize=3G
100,000,000	-Xms64G -Xmx64G -XX:NewSize=6G -XX:MaxNewSize=6G
200,000,000	-Xms96G -Xmx96G -XX:NewSize=9G -XX:MaxNewSize=9G
300,000,000	-Xms164G -Xmx164G -XX:NewSize=12G -XX:MaxNewSize=12G

9.13.85 ALM-14007 NameNode Heap Memory Usage Exceeds the Threshold

Description

The system checks the HDFS NameNode Heap Memory usage every 30 seconds and compares the actual Heap memory usage with the threshold. The HDFS NameNode Heap Memory usage has a default threshold. This alarm is generated when the HDFS NameNode Heap Memory usage exceeds the threshold.

You can change the threshold in **O&M** > **Alarm > Thresholds** > *Name of the desired cluster* **> HDFS**.

When the **Trigger Count** is 1, this alarm is cleared when the HDFS NameNode Heap memory usage is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the HDFS NameNode Heap memory usage is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14007	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

The HDFS NameNode Heap Memory usage is too high, which affects the data read/write performance of the HDFS.

Possible Causes

The HDFS NameNode Heap Memory is insufficient.

Procedure

Delete unnecessary files.

Step 1 Log in to the HDFS client as user **root**. Run **cd** to switch to the client installation directory, and run **source bigdata_env**.

If the cluster uses the security mode, perform security authentication.

Run the **kinit hdfs** command and enter the password as prompted. Obtain the password from the administrator.

Step 2 Run the **hdfs dfs -rm -r** *file or directory* command to delete unnecessary files.

Step 3 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check the NameNode JVM memory usage and configuration.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS**.

Step 5 In the **Basic Information** area, click **NameNode(Active)** to go to the HDFS WebUI.

NOTE

By default, the **admin** user does not have the permissions to manage other components. If the page cannot be opened or the displayed content is incomplete when you access the native UI of a component due to insufficient permissions, you can manually create a user with the permissions to manage that component.

Step 6 On the HDFS WebUI, click the **Overview** tab. In **Summary**, check the numbers of files, directories, and blocks in the HDFS.

Step 7 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS** > **Configurations** > **All Configurations**. In **Search**, enter **GC_OPTS** to check the **GC_OPTS** memory parameter of **HDFS->NameNode**.

Adjust the configuration in the system.

Step 8 Check whether the memory is configured properly based on the number of files in **Step 6** and the NameNode Heap Memory parameters in **Step 7**.

- If yes, go to **Step 9**.
- If no, go to **Step 11**.

NOTE

The recommended mapping between the number of HDFS file objects (filesystem objects = files + blocks) and the JVM parameters configured for NameNode is as follows:

- If the number of file objects reaches 10,000,000, you are advised to set the JVM parameters as follows: -Xms6G -Xmx6G -XX:NewSize=512M -XX:MaxNewSize=512M
- If the number of file objects reaches 20,000,000, you are advised to set the JVM parameters as follows: -Xms12G -Xmx12G -XX:NewSize=1G -XX:MaxNewSize=1G
- If the number of file objects reaches 50,000,000, you are advised to set the JVM parameters as follows: -Xms32G -Xmx32G -XX:NewSize=3G -XX:MaxNewSize=3G
- If the number of file objects reaches 100,000,000, you are advised to set the JVM parameters as follows: -Xms64G -Xmx64G -XX:NewSize=6G -XX:MaxNewSize=6G
- If the number of file objects reaches 200,000,000, you are advised to set the JVM parameters as follows: -Xms96G -Xmx96G -XX:NewSize=9G -XX:MaxNewSize=9G
- If the number of file objects reaches 300,000,000, you are advised to set the JVM parameters as follows: -Xms164G -Xmx164G -XX:NewSize=12G -XX:MaxNewSize=12G

Step 9 Modify the heap memory parameters of the NameNode based on the mapping between the number of file objects and the memory. Click **Save** and choose **Dashboard** > **More** > **Restart Service**.

Step 10 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Collect fault information.

Step 11 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 12 Select the following nodes in the required cluster from the **Service**:

- ZooKeeper
- HDFS

Step 13 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 14 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.86 ALM-14008 DataNode Heap Memory Usage Exceeds the Threshold

Description

The system checks the HDFS DataNode Heap Memory usage every 30 seconds and compares the actual Heap Memory usage with the threshold. The HDFS DataNode Heap Memory usage has a default threshold. This alarm is generated when the HDFS DataNode Heap Memory usage exceeds the threshold.

You can change the threshold in **O&M** > **Alarm > Thresholds** > *Name of the desired cluster* **> HDFS**.

When the **Trigger Count** is 1, this alarm is cleared when the HDFS DataNode Heap Memory usage is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the HDFS DataNode Heap Memory usage is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14008	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

The HDFS DataNode Heap Memory usage is too high, which affects the data read/write performance of the HDFS.

Possible Causes

The HDFS DataNode Heap Memory is insufficient.

Procedure

Delete unnecessary files.

Step 1 Log in to the HDFS client as user **root**. Run **cd** to switch to the client installation directory, and run **source bigdata_env**.

If the cluster uses the security mode, perform security authentication.

Run the **kinit hdfs** command and enter the password as prompted. Obtain the password from the administrator.

Step 2 Run the **hdfs dfs -rm -r** *file or directory* command to delete unnecessary files.

Step 3 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check the DataNode JVM memory usage and configuration.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS**.

Step 5 In the **Basic Information** area, click **NameNode(Active)** to go to the HDFS WebUI.

NOTE

By default, the **admin** user does not have the permissions to manage other components. If the page cannot be opened or the displayed content is incomplete when you access the native UI of a component due to insufficient permissions, you can manually create a user with the permissions to manage that component.

Step 6 On the HDFS WebUI, click the **DataNodes** tab, and check the number of blocks of all DataNodes related to the alarm.

Step 7 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS** > **Configurations** > **All Configurations**. In **Search**, enter **GC_OPTS** to check the GC_OPTS memory parameter of **HDFS->DataNode**.

Adjust the configuration in the system.

Step 8 Check whether the memory is configured properly based on the number of block in **Step 6** and the DataNode Heap Memory parameters in **Step 7**.

- If yes, go to **Step 9**.
- If no, go to **Step 11**.

NOTE

The mapping between the average number of blocks of a DataNode instance and the DataNode memory is as follows:

- If the average number of blocks of a DataNode instance reaches 2,000,000, the reference values of the JVM parameters of the DataNode are as follows: -Xms6G -Xmx6G -XX:NewSize=512M -XX:MaxNewSize=512M
- If the average number of blocks of a DataNode instance reaches 5,000,000, the reference values of the JVM parameters of the DataNode are as follows: -Xms12G -Xmx12G -XX:NewSize=1G -XX:MaxNewSize=1G

Step 9 Modify the heap memory parameters of the DataNode based on the mapping between the number of blocks and the memory. Click **Save** and choose **Dashboard** > **More** > **Restart Service**.

Step 10 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Collect fault information.

Step 11 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 12 Select **HDFS** in the required cluster from the **Service**.

Step 13 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 14 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.87 ALM-14009 Number of Dead DataNodes Exceeds the Threshold

Description

The system periodically detects the number of dead DataNodes in the HDFS cluster every 30 seconds, and compares the number with the threshold. The number of DataNodes in the Dead state has a default threshold. This alarm is generated when the number exceeds the threshold.

You can change the threshold in **O&M** > **Alarm > Thresholds** > *Name of the desired cluster* > **HDFS**.

When the **Trigger Count** is 1, this alarm is cleared when the number of Dead DataNodes is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the number of Dead DataNodes is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14009	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
NameServiceName	Specifies the NameService for which the alarm is generated.
Trigger condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

DataNodes that are in the Dead state cannot provide HDFS services.

Possible Causes

- DataNodes are faulty or overloaded.
- The network between the NameNode and the DataNode is disconnected or busy.
- NameNodes are overloaded.
- The NameNodes are not restarted after the DataNode is deleted.

Procedure

Check whether DataNodes are faulty.

Step 1 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS**. The **HDFS Status** page is displayed.

Step 2 In the **Basic Information** area, click **NameNode(Active)** to go to the HDFS WebUI.

NOTE

By default, the **admin** user does not have the permissions to manage other components. If the page cannot be opened or the displayed content is incomplete when you access the native UI of a component due to insufficient permissions, you can manually create a user with the permissions to manage that component.

Step 3 On the HDFS WebUI, click the **Datanodes** tab. In the **In operation** area, click **Filter** to check whether **down** is in the drop-down list.

- If yes, select **down**, record the information about the filtered DataNodes, and go to **Step 4**.
- If no, go to **Step 8**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* > **Services** > **HDFS** > **Instance** to check whether recorded DataNodes exist in the instance list.

- If all recorded DataNodes exist, go to **Step 5**.
- If none of the recorded DataNodes exists, go to **Step 6**.
- If some of the recorded DataNodes exist, go to **Step 7**.

Step 5 Locate the DataNode instance, click **More** > **Restart Instance** to restart it and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Step 6 Select all NameNode instances, choose **More** > **Instance Rolling Restart** to restart them and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 16**.

Step 7 Select all NameNode instances, choose **More** > **Instance Rolling Restart** to restart them. Locate the DataNode instance, click **More** > **Restart Instance** to restart it and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Check the status of the network between the NameNode and the DataNode.

Step 8 Log in to the faulty DataNode on the management page as user **root**, and run the **ping** *IP address of the NameNode* command to check whether the network between the DataNode and the NameNode is abnormal.

On the FusionInsight Manager page, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Instance**. In the instance list, view the service plane IP address of the faulty DataNode.

- If yes, go to **Step 9**.
- If no, go to **Step 10**.

Step 9 Rectify the network fault, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 10**.

Check whether the DataNode is overloaded.

Step 10 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms** and check whether the alarm **ALM-14008 HDFS DataNode Memory Usage Exceeds the Threshold** exists.

- If yes, go to **Step 11**.
- If no, go to **Step 13**.

Step 11 See **ALM-14008 HDFS DataNode Memory Usage Exceeds the Threshold** to handle the alarm and check whether the alarm is cleared.

- If yes, go to **Step 12**.
- If no, go to **Step 13**.

Step 12 Check whether the alarm is cleared from the alarm list.

- If yes, no further action is required.
- If no, go to **Step 13**.

Check whether the NameNode is overloaded.

Step 13 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms** and check whether the alarm **ALM-14007 HDFS NameNode Memory Usage Exceeds the Threshold** exists.

- If yes, go to **Step 14**.
- If no, go to **Step 16**.

Step 14 See **ALM-14007 HDFS NameNode Memory Usage Exceeds the Threshold** to handle the alarm and check whether the alarm is cleared.

- If yes, go to **Step 15**.
- If no, go to **Step 16**.

Step 15 Check whether the alarm is cleared from the alarm list.

- If yes, no further action is required.
- If no, go to **Step 16**.

Collect fault information.

Step 16 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 17 Select **HDFS** in the required cluster from the **Service**.

Step 18 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 19 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.88 ALM-14010 NameService Service Is Abnormal

Description

The system checks the NameService service status every 180 seconds. This alarm is generated when the NameService service is unavailable.

This alarm is cleared when the NameService service recovers.

Attribute

Alarm ID	Alarm Severity	Auto Clear
14010	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
NameServiceName	Specifies the NameService for which the alarm is generated.

Impact on the System

HDFS fails to provide services for upper-layer components based on the NameService service, such as HBase and MapReduce. As a result, users cannot read or write files.

Possible Causes

- The KrbServer service is abnormal.
- The JournalNode is faulty.
- The DataNode is faulty.
- The disk capacity is insufficient.

- The NameNode enters safe mode.

Procedure

Check the KrbServer service status.

Step 1 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 2 Check whether the KrbServer service exists.

- If yes, go to **Step 3**.
- If no, go to **Step 6**.

Step 3 Click **KrbServer**.

Step 4 Click **Instances**. On the KrbServer management page, select the faulty instance, and choose **More** > **Restart Instance**. Check whether the instance successfully restarts.

- If yes, go to **Step 5**.
- If no, go to **Step 24**.

Step 5 Choose **O&M** > **Alarm** > **Alarms** and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Check the JournalNode instance status.

Step 6 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 7 Choose **HDFS** > **Instances**.

Step 8 Check whether the **Running Status** of the JournalNode is **Normal**.

- If yes, go to **Step 11**.
- If no, go to **Step 9**.

Step 9 Select the faulty JournalNode, and choose **More** > **Restart Instance**. Check whether the JournalNode successfully restarts.

- If yes, go to **Step 10**.
- If no, go to **Step 24**.

Step 10 Choose **O&M** > **Alarm** > **Alarms** and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Check the DataNode instance status.

Step 11 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS**.

Step 12 Click **Instances** and check whether **Running Status** of all DataNodes is **Normal**.

- If yes, go to **Step 15**.
- If no, go to **Step 13**.

Step 13 Click **Instances**. On the DataNode management page, select the faulty instance, and choose **More** > **Restart Instance**. Check whether the DataNode successfully restarts.

- If yes, go to **Step 14**.
- If no, go to **Step 15**.

Step 14 Choose **O&M** > **Alarm** > **Alarms** and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 15**.

Check disk status.

Step 15 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Host**.

Step 16 In the **Disk** column, check whether the disk space is insufficient.

- If yes, go to **Step 17**.
- If no, go to **Step 19**.

Step 17 Expand the disk capacity.

Step 18 Choose **O&M** > **Alarm** > **Alarms** and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 19**.

Check whether NameNode is in the safe mode.

Step 19 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS**. Click **NameNode(Active)** of the abnormal NameService. The NameNode web UI is displayed.

NOTE

By default, the admin user does not have the management rights of other components. If the page cannot be opened or the content is not completely displayed due to insufficient permission when you access the native page of a component, you can manually create a user with the management rights of the corresponding component to log in to the component.

Step 20 On the NameNode web UI, check whether "Safe mode is ON." is displayed.

Information behind **Safe mode is ON** is alarm information and is displayed based actual conditions.

- If yes, go to **Step 21**.
- If no, go to **Step 24**.

Step 21 Log in to the client as user **root**. Run the **cd** command to go to the client installation directory and run the **source bigdata_env** command. If the cluster uses the security mode, perform security authentication. Run the **kinit hdfs** command and enter the password as prompted. The password can be obtained from the MRS cluster administrator. If the cluster uses the non-security mode, log in as user **omm** and run the command. Ensure that user **omm** has the client execution permission.

Step 22 Run **hdfs dfsadmin -safemode leave**.

Step 23 Choose **O&M** > **Alarm** > **Alarms** and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 24**.

Collect the fault information.

Step 24 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 25 In the **Service** area, select the following nodes of the desired cluster.

- ZooKeeper
- HDFS

Step 26 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 27 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.89 ALM-14011 DataNode Data Directory Is Not Configured Properly

Description

The DataNode parameter **dfs.datanode.data.dir** specifies DataNode data directories. This alarm is generated when a configured data directory cannot be created, a data directory uses the same disk as other critical directories in the system, or multiple directories use the same disk immediately.

This alarm is cleared when the DataNode data directory is configured properly and this DataNode for which the alarm is generated is restarted.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14011	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

If the DataNode data directory is mounted to the root directory or a critical directory, the disk space of the root directory or critical directory will be used up after long time running and the system will be faulty.

If the DataNode data directory is not configured properly, HDFS performance will deteriorate.

Possible Causes

- The DataNode data directory fails to be created.
- The DataNode data directory uses the same disk with critical directories, such as **/** or **/boot**.
- Multiple directories in the DataNode data directory use the same disk.

Procedure

Check the alarm cause and information about the DataNode for which the alarm is generated.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms**. In the alarm list, click the alarm.

Step 2 In **HostName** of **Location**, obtain the host name of the DataNode for which the alarm is generated.

Delete directories that do not comply with the disk plan from the DataNode data directory.

Step 3 Choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Instance**. In the instance list, click the DataNode instance on the node for which the alarm is generated.

Step 4 Click **Instance Configurations** and view the value of the DataNode parameter **dfs.datanode.data.dir**.

Step 5 Check whether all DataNode data directories are consistent with the disk plan.

- If yes, go to **Step 6**.
- If no, go to **Step 9**.

Step 6 Modify the DataNode parameter **dfs.datanode.data.dir** and delete the incorrect directories.

Step 7 Choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS** > **Instance** and restart the DataNode instance.

Step 8 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Step 9 Log in to the DataNode for which the alarm is generated as user **root**.

- If the alarm cause is "The DataNode data directory fails to be created", go to **Step 10**.
- If the alarm cause is "The DataNode data directory uses the same disk with critical directories, such **/** or **/boot**", go to **Step 17**.
- If the alarm cause is "Multiple directories in the DataNode data directory uses the same disk", go to **Step 21**.

Check whether the DataNode data directory fails to be created.

Step 10 Run the **su - omm** command to switch to user **omm**.

Step 11 Run the **ls** command to check whether the directories exist in the DataNode data directory.

- If yes, go to **Step 26**.
- If no, go to **Step 12**.

Step 12 Run the **mkdir** *data directory* command to create the directory and check whether the directory can be successfully created.

- If yes, go to **Step 24**.
- If no, go to **Step 13**.

Step 13 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms** to check whether alarm **ALM-12017 Insufficient Disk Capacity** exists.

- If yes, go to **Step 14**.
- If no, go to **Step 15**.

Step 14 Adjust the disk capacity and check whether alarm **ALM-12017 Insufficient Disk Capacity** is cleared. For details, see **ALM-12017 Insufficient Disk Capacity**.

- If yes, go to **Step 12**.
- If no, go to **Step 15**.

Step 15 Check whether user **omm** has the **rwx** or **x** permission of all the upper-layer directories of the directory. (For example, for **/tmp/abc/**, user **omm** has the **x** permission for directory **tmp** and the **rwx** permission for directory **abc**.)

- If yes, go to **Step 24**.
- If no, go to **Step 16**.

Step 16 Run the **chmod u+rwx** *path* or **chmod u+x** *path* command as user **root** to assign the **rwx** or **x** permission of these directories to user **omm**. Then go to **Step 12**.

Check whether the DataNode data directory use the same disk as other critical directories in the system.

Step 17 Run the **df** command to obtain the disk mounting information of each directory in the DataNode data directory.

Step 18 Check whether the directories mounted to the disk are critical directories, such as **/** or **/boot**.

- If yes, go to **Step 19**.
- If no, go to **Step 24**.

Step 19 Change the value of the DataNode parameter **dfs.datanode.data.dir** and delete the directories that use the same disk as critical directories.

Step 20 Go to **Step 24**.

Check whether multiple directories in the DataNode data directory use the same disk.

Step 21 Run the **df** command to obtain the disk mounting information of each directory in the DataNode data directory. Record the mounted directory in the command output.

Step 22 Modify the DataNode node parameters **dfs.datanode.data.dir** to reserve only one directory among the directories that mounted to the same disk directory.

Step 23 Go to **Step 24**.

Restart the DataNode and check whether the alarm is cleared.

Step 24 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Instance** and restart the DataNode instance

Step 25 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 26**.

Collect fault information.

Step 26 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 27 Select **HDFS** in the required cluster from the **Service**.

Step 28 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 29 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.90 ALM-14012 JournalNode Is Out of Synchronization

Description

On the active NameNode, the system checks the data consistency of all JournalNodes in the cluster every 5 minutes. This alarm is generated when the data on a JournalNode is inconsistent with the data on the other JournalNodes.

This alarm is cleared in 5 minutes after the data on JournalNodes is consistent.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14012	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
NameServiceName	Specifies the NameService for which the alarm is generated.

Impact on the System

When a JournalNode is working incorrectly, the data on the node becomes inconsistent with that on the other JournalNodes. If data on more than half of JournalNodes is inconsistent, the NameNode cannot work correctly, making the HDFS service unavailable.

Possible Causes

- The JournalNode instance does not exist (deleted or migrated).
- The JournalNode instance has not been started or has been stopped.
- The JournalNode instance is working incorrectly.
- The network of the JournalNode is unreachable.

Procedure

Check whether the JournalNode instance has been started up.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms**. In the alarm list, click the alarm.

Step 2 Check **Location** and obtain the IP address of the JournalNode for which the alarm is generated.

Step 3 Choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Instance**. In the instance list, check whether the JournalNode instance exists on the node for which the alarm is generated.

- If yes, go to **Step 5**.
- If no, go to **Step 4**.

Step 4 Choose **O&M** > **Alarm** > **Alarms**. In the alarm list, click **Clear** in the **Operation** column of the alarm. In the dialog box that is displayed, click **OK**. No further action is needed.

Step 5 Click the JournalNode instance and check whether its **Configuration Status** is **Synchronized**.

- If yes, go to **Step 8**.
- If no, go to **Step 6**.

Step 6 Select the JournalNode instance and choose **Start Instance** to start the instance.

Step 7 After 5 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 15**.

Check whether the JournalNode instance is working correctly.

Step 8 Check whether **Running Status** of the JournalNode instance is **Normal**.

- If yes, go to **Step 11**.
- If no, go to **Step 9**.

Step 9 Select the JournalNode instance and choose **More** > **Restart Instance** to start the instance.

Step 10 After 5 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 15**.

Check whether the network of the JournalNode is reachable.

Step 11 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Instance** to check the service IP address of the active NameNode.

Step 12 Log in to the active NameNode as user **root**.

Step 13 Run the **ping** command to check whether a timeout occurs or the network is unreachable between the active NameNode and the JournalNode.

ping *service IP address of the JournalNode*

- If yes, go to **Step 14**.
- If no, go to **Step 15**.

Step 14 Contact the network administrator to rectify the network fault and check whether the alarm is cleared 5 minutes later.

- If yes, no further action is required.
- If no, go to **Step 15**.

Collect fault information.

Step 15 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 16 Select **HDFS** in the required cluster from the **Service**.

Step 17 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 18 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.91 ALM-14013 Failed to Update the NameNode FsImage File

Description

HDFS metadata is stored in the FsImage file of the NameNode data directory, which is specified by the **dfs.namenode.name.dir** configuration item. The standby NameNode periodically combines existing FsImage files and Editlog files stored in the JournalNode to generate a new FsImage file, and then pushes the new FsImage file to the data directory of the active NameNode. This period is specified by the **dfs.namenode.checkpoint.period** configuration item of HDFS. The default value is 3600s, namely, one hour. If the FsImage file in the data directory of the active NameNode is not updated, the HDFS metadata combination function is abnormal and requires rectification.

On the active NameNode, the system checks the FsImage file information every five minutes. This alarm is generated when no FsImage file is generated within three combination periods.

This alarm is cleared when a new FsImage file is generated and pushed to the active NameNode, which indicates that the HDFS metadata combination function can be properly used.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14013	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
NameServiceName	Specifies the NameService for which the alarm is generated.

Impact on the System

If the FsImage file in the data directory of the active NameNode is not updated, the HDFS metadata combination function is abnormal and requires rectification. If it is not rectified, the Editlog files increase continuously after HDFS runs for a period. In this case, HDFS restart is time-consuming because a large number of Editlog files need to be loaded. In addition, this alarm also indicates that the standby NameNode is abnormal and the NameNode high availability (HA) mechanism becomes invalid. When the active NameNode is faulty, the HDFS service becomes unavailable.

Possible Causes

- The standby NameNode is stopped.
- The standby NameNode instance is working incorrectly.
- The standby NameNode fails to generate a new FsImage file.
- Space of the data directory on the standby NameNode is insufficient.
- The standby NameNode fails to push the FsImage file to the active NameNode.
- Space of the data directory on the active NameNode is insufficient.

Procedure

Check whether the standby NameNode is stopped.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms**. In the alarm list, click the alarm.

Step 2 View **Location** and obtain the host name of the active NameNode for which the alarm is generated and name of the NameService where the active NameNode resides.

Step 3 Choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Instance**, find the standby NameNode instance of the NameService in the instance list, and check whether its **Configuration Status** is **Synchronized**.

- If yes, go to **Step 6**.
- If no, go to **Step 4**.

Step 4 Select the standby NameNode instance, choose **Start Instance**, and wait until the startup is complete.

Step 5 Wait for a NameNode metadata combination period and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Check whether the NameNode instance is working correctly.

Step 6 Check whether **Running Status** of the standby NameNode instance is **Normal**.

- If yes, go to **Step 9**.
- If no, go to **Step 7**.

Step 7 Select the standby NameNode instance, choose **More** > **Restart Instance**, and wait until the startup is complete.

Step 8 Wait for a NameNode metadata combination period and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 30**.

Check whether the standby NameNode fails to generate a new FsImage file.

Step 9 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Configurations** > **All Configurations**, and search and obtain the value of **dfs.namenode.checkpoint.period**. This value is the period of NameNode metadata combination.

Step 10 Choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Instance** and obtain the service IP addresses of the active and standby NameNodes of the NameService for which the alarm is generated.

Step 11 Click the **NameNode(***xx***,Standy)** and **Instance Configurations** to obtain the value of **dfs.namenode.name.dir**. This value is the FsImage storage directory of the standby NameNode.

Step 12 Log in to the standby NameNode as user **root** or **omm**.

Step 13 Go to the FsImage storage directory and check the generation time of the newest FsImage file.

cd *Storage directory of the standby NameNode***/current**

```
stat -c %y $(ls -t | grep "fsimage_[0-9]*$" | head -1)
```

Step 14 Run the **date** command to obtain the current system time.

Step 15 Calculate the time difference between the generation time of the newest FsImage file and the current system time and check whether the time difference is greater than three times of the metadata combination period.

- If yes, go to **Step 16**.
- If no, go to **Step 20**.

Step 16 The metadata combination function of the standby NameNode is faulty. Run the following command to check whether the fault is caused by insufficient storage space.

Go to the FsImage storage directory and check the size of the newest FsImage file (in MB).

```
cd Storage directory of the standby NameNode/current
```

```
du -m $(ls -t | grep "fsimage_[0-9]*$" | head -1) | awk '{print $1}'
```

Step 17 Run the following command to check the available disk space of the standby NameNode (in MB).

```
df -m ./ | awk 'END{print $4}'
```

Step 18 Compare the FsImage file size and the available disk space and determine whether another FsImage file can be stored on the disk.

- If yes, go to **Step 7**.
- If no, go to **Step 19**.

Step 19 Clear the redundant files on the disk where the directory resides to reserve sufficient space for metadata. After the clearance, wait for a NameNode metadata combination period and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 20**.

Check whether the standby NameNode fails to push the FsImage file to the active NameNode.

Step 20 Log in to the standby NameNode as user **root**.

Step 21 Run the **su - omm** command to switch to user **omm**.

Step 22 Run the following command to check whether the standby NameNode can push the file to the active NameNode.

```
tmpFile=/tmp/tmp_test_$(date +%s)
```

```
echo "test" > $tmpFile
```

```
scp $tmpFile Service IP address of the active NameNode:/tmp
```

- If yes, go to **Step 24**.
- If no, go to **Step 23**.

Step 23 When the standby NameNode fails to push data to the active NameNode as user **omm**, contact the system administrator to handle the fault. Wait for a NameNode metadata combination period and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 24**.

Check whether space on the data directory of the active NameNode is insufficient.

Step 24 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Instance**, click the active NameNode of the NameService for which the alarm is generated, and then click **Instance Configurations** to obtain the value of **dfs.namenode.name.dir**. This value is the FsImage storage directory of the active NameNode.

Step 25 Log in to the active NameNode as user **root** or **omm**.

Step 26 Go to the FsImage storage directory and check the size of the newest FsImage file (in MB).

```
cd Storage directory of the active NameNode/current
```

```
du -m $(ls -t | grep "fsimage_[0-9]*$" | head -1) | awk '{print $1}'
```

Step 27 Run the following command to check the available disk space of the active NameNode (in MB).

```
df -m ./ | awk 'END{print $4}'
```

Step 28 Compare the FsImage file size and the available disk space and determine whether another FsImage file can be stored on the disk.

- If yes, go to **Step 30**.
- If no, go to **Step 29**.

Step 29 Clear the redundant files on the disk where the directory resides to reserve sufficient space for metadata. After the clearance, wait for a NameNode metadata combination period and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 30**.

Collect fault information.

Step 30 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 31 Select **NameNode** in the required cluster from the **Service**.

Step 32 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 33 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.92 ALM-14014 NameNode GC Time Exceeds the Threshold

Description

The system checks the garbage collection (GC) duration of the NameNode process every 60 seconds. This alarm is generated when the GC duration exceeds the threshold (12 seconds by default).

This alarm is cleared when the GC duration is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14014	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

A long GC duration of the NameNode process may interrupt the services.

Possible Causes

The heap memory of the NameNode instance is overused or the heap memory is inappropriately allocated. As a result, GCs occur frequently.

Procedure

Check the GC duration.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms**. On the displayed interface, click the drop-down button of **ALM-14014 NameNode GC Time Exceeds the Threshold.** Then check the role name in **Location** and confirm the IP adress of the instance.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Instance** > **NameNode (IP address for which the alarm is generated)**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **Garbage Collection**, and select **NameNode Garbage Collection (GC)** to check the GC duration statistics of the NameNode process collected every minute.

Step 3 Check whether the GC duration of the NameNode process collected every minute exceeds the threshold (12 seconds by default).

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Configurations** > **All Configurations** > **NameNode** > **System** to increase the value of **GC_OPTS** parameter as required.

NOTE

The recommended mapping between the number of HDFS file objects (filesystem objects = files + blocks) and the JVM parameters configured for NameNode is as follows:

- If the number of file objects reaches 10,000,000, you are advised to set the JVM parameters as follows: -Xms6G -Xmx6G -XX:NewSize=512M -XX:MaxNewSize=512M
- If the number of file objects reaches 20,000,000, you are advised to set the JVM parameters as follows: -Xms12G -Xmx12G -XX:NewSize=1G -XX:MaxNewSize=1G
- If the number of file objects reaches 50,000,000, you are advised to set the JVM parameters as follows: -Xms32G -Xmx32G -XX:NewSize=3G -XX:MaxNewSize=3G
- If the number of file objects reaches 100,000,000, you are advised to set the JVM parameters as follows: -Xms64G -Xmx64G -XX:NewSize=6G -XX:MaxNewSize=6G
- If the number of file objects reaches 200,000,000, you are advised to set the JVM parameters as follows: -Xms96G -Xmx96G -XX:NewSize=9G -XX:MaxNewSize=9G
- If the number of file objects reaches 300,000,000, you are advised to set the JVM parameters as follows: -Xms164G -Xmx164G -XX:NewSize=12G -XX:MaxNewSize=12G

Step 5 Save the configuration and restart the NameNode instance.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **NameNode** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.93 ALM-14015 DataNode GC Time Exceeds the Threshold

Description

The system checks the garbage collection (GC) duration of the DataNode process every 60 seconds. This alarm is generated when the GC duration exceeds the threshold (12 seconds by default).

This alarm is cleared when the GC duration is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14015	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

A long GC duration of the DataNode process may interrupt the services.

Possible Causes

The heap memory of the DataNode instance is overused or the heap memory is inappropriately allocated. As a result, GCs occur frequently.

Procedure

Check the GC duration.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms**. On the displayed interface, click the drop-down button of **ALM-14015 DataNode GC Time Exceeds the Threshold**. Then check the role name in **Location** and confirm the IP adress of the instance.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Instance** > **DataNode (IP address for which the alarm is generated)**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **Garbage Collection**, and select **DataNode Garbage Collection (GC)** to check the GC duration statistics of the DataNode process collected every minute.

Step 3 Check whether the GC duration of the DataNode process collected every minute exceeds the threshold (12 seconds by default).

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Configurations** > **All Configurations** > **DataNode** > **System** to increase the value of **GC_OPTS** parameter as required.

NOTE

The mapping between the average number of blocks of a DataNode instance and the DataNode memory is as follows:

- If the average number of blocks of a DataNode instance reaches 2,000,000, the reference values of the JVM parameters of the DataNode are as follows: -Xms6G -Xmx6G -XX:NewSize=512M -XX:MaxNewSize=512M
- If the average number of blocks of a DataNode instance reaches 5,000,000, the reference values of the JVM parameters of the DataNode are as follows: -Xms12G -Xmx12G -XX:NewSize=1G -XX:MaxNewSize=1G

Step 5 Save the configuration and restart the DataNode instance.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **DataNode** in the required cluster from the **Service**.

Step 9 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.94 ALM-14016 DataNode Direct Memory Usage Exceeds the Threshold

Description

The system checks the direct memory usage of HDFS every 30 seconds. This alarm is generated when the direct memory usage of DataNode instances exceeds the threshold (90% of the maximum memory).

This alarm is automatically cleared when the direct memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
14016	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

If the available direct memory of DataNode instances is insufficient, a memory overflow may occur and the service breaks down.

Possible Causes

The direct memory of DataNode instances is overused or the direct memory is inappropriately allocated.

Procedure

Check the direct memory usage.

Step 1 On the **Home** page of FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. On the page that is displayed, click the drop-down list in the row containing **ALM-14016 DataNode Direct Memory Usage Exceeds the Threshold**, and view the role name and IP address of the instance for which the alarm is generated in the **Location** area.

Step 2 On the **Home** page of FusionInsight Manager, choose **Cluster** > **Services** > **HDFS**. On the page that is displayed, click the **Instance** tab. In the instance list, select **DataNode** (IP address of the instance for which this alarm is generated). Click the drop-down list in the upper right corner of the chart, choose **Customize** > **Resource**, and select **DataNode Memory** to check the direct memory usage.

Step 3 Check whether the used direct memory of a DataNode instance reaches 90% (default threshold) of the maximum direct memory allocated to it.

- If yes, go to **Step 4**.
- If no, go to **Step 8**.

Step 4 On the **Home** page of FusionInsight Manager, choose **Cluster** > **Services** > **HDFS**. On the page that is displayed, click the **Configuration** tab then the **All Configurations** sub-tab, and select **DataNode** > **System**. Check whether **-XX:MaxDirectMemorySize** exists in the **GC_OPTS** parameter.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 Adjust the value of **-XX:MaxDirectMemorySize**.

1. In **GC_OPTS**, check the value of **-Xmx** and check whether the node memory is sufficient.

> **NOTE**
>
> You can determine whether the node memory is sufficient based on the actual environment. For example, you can use the following method:
>
> Use the IP address to log in to the instance for which the alarm is generated as user **root** and run the **free -g** command to check the value of **Mem** in the **free** column. The value indicates the available memory of the node. In the following example, the available memory of the node is 4 GB.
>
> ```
>              total        used        free      shared  buff/cache   available
> Mem:           112          48           4          10          58          46
> ......
> ```
>
> If the value of **Mem** is at least that of **-Xmx**, the node memory is sufficient. If the value of **Mem** is less than that of **-Xmx**, the node memory is insufficient.

– If yes, change the value of **-XX:MaxDirectMemorySize** to that of **-Xmx**.

– If no, increase **-XX:MaxDirectMemorySize** to a value no larger than that of **Mem**.

2. Save the configuration and restart the DataNode instances.

Step 6 Check whether **ALM-14008 DataNode Heap Memory Usage Exceeds the Threshold** exists.

- If yes, rectify the fault by referring to **ALM-14008 DataNode Heap Memory Usage Exceeds the Threshold**.
- If no, go to **Step 7**.

Step 7 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Collect the fault information.

Step 8 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 9 Expand the **Service** drop-down list, and select **DataNode** for the target cluster.

Step 10 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.95 ALM-14017 NameNode Direct Memory Usage Exceeds the Threshold

Description

The system checks the direct memory usage of the HDFS service every 30 seconds. This alarm is generated when the direct memory usage of a NameNode instance exceeds the threshold (90% of the maximum memory).

The alarm is cleared when the direct memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14017	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available direct memory of the HDFS service is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The direct memory of the NameNode instance is overused or the direct memory is inappropriately allocated.

Procedure

Check the direct memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms.** On the displayed interface, click the drop-down button of **ALM-14017 NameNode Direct Memory Usage Exceeds the Threshold**. Then check the role name in **Location** and confirm the IP adress of the instance.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Instance** > **NameNode (IP address for which the alarm is generated)**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **Resource**, and select **NameNode Memory** to check the direct memory usage.

Step 3 Check whether the used direct memory of NameNode reaches 90% of the maximum direct memory specified for NameNode by default.

- If yes, go to **Step 4**.
- If no, go to **Step 8**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Configurations** > **All Configurations** > **NameNode** > **System** to check whether "-XX:MaxDirectMemorySize" exists in the **GC_OPTS** parameter.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 In the **GC_OPTS** parameter, delete "-XX:MaxDirectMemorySize". Save the configuration and restart the NameNode instance.

Step 6 Check whether the **ALM-14007 NameNode Heap Memory Usage Exceeds the Threshold** exists.

- If yes, handle the alarm by referring to **ALM-14007 NameNode Heap Memory Usage Exceeds the Threshold**.
- If no, go to **Step 7**.

Step 7 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Collect fault information.

Step 8 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 9 Select **NameNode** in the required cluster from the **Service**.

Step 10 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.96 ALM-14018 NameNode Non-heap Memory Usage Exceeds the Threshold

Description

The system checks the non-heap memory usage of the HDFS NameNode every 30 seconds and compares the actual usage with the threshold. The non-heap memory usage of the HDFS NameNode has a default threshold. This alarm is generated when the non-heap memory usage of the HDFS NameNode exceeds the threshold.

Users can choose **O&M > Alarm > Thresholds >** *Name of the desired cluster* > **HDFS** to change the threshold.

This alarm is cleared when the no-heap memory usage of the HDFS NameNode is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14018	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the memory usage of the HDFS NameNode is too high, data read/write performance of HDFS will be affected.

Possible Causes

Non-heap memory of the HDFS NameNode is insufficient.

Procedure

Delete unnecessary files.

Step 1 Log in to the HDFS client as user **root**. Run the **cd** command to go to the client installation directory, and run the **source bigdata_env** command.

If the cluster adopts the security mode, perform security authentication.

Run the **kinit hdfs** command and enter the password as prompted. Obtain the password from the administrator.

Step 2 Run the **hdfs dfs -rm -r** *file or directory path* command to delete unnecessary files.

Step 3 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check the NameNode JVM non-heap memory usage and configuration.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS**. The HDFS status page is displayed.

Step 5 In the **Basic Information** area, click **NameNode(Active)**. The HDFS WebUI is displayed.

NOTE

By default, the **admin** user does not have the permissions to manage other components. If the page cannot be opened or the displayed content is incomplete when you access the native UI of a component due to insufficient permissions, you can manually create a user with the permissions to manage that component.

Step 6 On the HDFS WebUI, click the **Overview** tab. In **Summary**, check the numbers of files, directories, and blocks in HDFS.

Step 7 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS** > **Configurations** > **All Configurations**. In **Search**, enter **GC_OPTS** to check the **GC_OPTS** non-heap memory parameter of **HDFS->NameNode**.

Adjust system configurations.

Step 8 Check whether the non-heap memory is properly configured based on the number of file objects in **Step 6** and the non-heap parameters configured for NameNode in **Step 7**.

- If yes, go to **Step 9**.

- If no, go to **Step 12**.

NOTE

The recommended mapping between the number of HDFS file objects (filesystem objects = files + blocks) and the JVM parameters configured for NameNode is as follows:

- If the number of file objects reaches 10,000,000, you are advised to set the JVM parameters as follows: -Xms6G -Xmx6G -XX:NewSize=512M -XX:MaxNewSize=512M
- If the number of file objects reaches 20,000,000, you are advised to set the JVM parameters as follows: -Xms12G -Xmx12G -XX:NewSize=1G -XX:MaxNewSize=1G
- If the number of file objects reaches 50,000,000, you are advised to set the JVM parameters as follows: -Xms32G -Xmx32G -XX:NewSize=3G -XX:MaxNewSize=3G
- If the number of file objects reaches 100,000,000, you are advised to set the JVM parameters as follows: -Xms64G -Xmx64G -XX:NewSize=6G -XX:MaxNewSize=6G
- If the number of file objects reaches 200,000,000, you are advised to set the JVM parameters as follows: -Xms96G -Xmx96G -XX:NewSize=9G -XX:MaxNewSize=9G
- If the number of file objects reaches 300,000,000, you are advised to set the JVM parameters as follows: -Xms164G -Xmx164G -XX:NewSize=12G -XX:MaxNewSize=12G

Step 9 Modify the **GC_OPTS** parameter of the NameNode based on the mapping between the number of file objects and non-heap memory.

Step 10 Save the configuration and click **Dashboard** > **More** > **Restart Service**.

Step 11 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 12**.

Collect fault information.

Step 12 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 13 Select the following services in the required cluster from the **Service**.

- ZooKeeper
- HDFS

Step 14 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 15 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.97 ALM-14019 DataNode Non-heap Memory Usage Exceeds the Threshold

Description

The system checks the non-heap memory usage of the HDFS DataNode every 30 seconds and compares the actual usage with the threshold. The non-heap memory usage of the HDFS DataNode has a default threshold. This alarm is generated when the non-heap memory usage of the HDFS DataNode exceeds the threshold.

Users can choose **O&M > Alarm > Thresholds>** *Name of the desired cluster* **> HDFS** to change the threshold.

This alarm is cleared when the no-heap memory usage of the HDFS DataNode is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14019	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the memory usage of the HDFS DataNode is too high, data read/write performance of HDFS will be affected.

Possible Causes

Non-heap memory of the HDFS DataNode is insufficient.

Procedure

Delete unnecessary files.

Step 1 Log in to the HDFS client as user **root**. Run the **cd** command to go to the client installation directory, and run the **source bigdata_env** command.

If the cluster adopts the security mode, perform security authentication.

Run the **kinit hdfs** command and enter the password as prompted. Obtain the password from the administrator.

Step 2 Run the **hdfs dfs -rm -r** *file or directory path* command to delete unnecessary files.

Step 3 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check the DataNode JVM non-heap memory usage and configuration.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS**.

Step 5 In the **Basic Information** area, click **NameNode(Active)**. The HDFS WebUI is displayed.

NOTE

By default, the **admin** user does not have the permissions to manage other components. If the page cannot be opened or the displayed content is incomplete when you access the native UI of a component due to insufficient permissions, you can manually create a user with the permissions to manage that component.

Step 6 On the HDFS WebUI, click the **Datanodes** tab to view the number of blocks of all DataNodes that report alarms.

Step 7 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Configurations** > **All Configurations**. In **Search**, enter **GC_OPTS** to check the **GC_OPTS** non-heap memory parameter of **HDFS->DataNode**.

Adjust system configurations.

Step 8 Check whether the memory is properly configured based on the number of blocks in **Step 6** and the memory parameters configured for DataNode in **Step 7**.

- If yes, go to **Step 9**.
- If no, go to **Step 12**.

NOTE

The mapping between the average number of blocks of a DataNode instance and the DataNode memory is as follows:

- If the average number of blocks of a DataNode instance reaches 2,000,000, the reference values of the JVM parameters of the DataNode are as follows: -Xms6G -Xmx6G -XX:NewSize=512M -XX:MaxNewSize=512M
- If the average number of blocks of a DataNode instance reaches 5,000,000, the reference values of the JVM parameters of the DataNode are as follows: -Xms12G -Xmx12G -XX:NewSize=1G -XX:MaxNewSize=1G

Step 9 Modify the **GC_OPTS** parameter of the DataNode based on the mapping between the number of blocks and memory.

Step 10 Save the configuration and click **Dashboard** > **More** > **Restart Service**.

Step 11 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 12**.

Collect fault information.

Step 12 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 13 Select the following services in the required cluster from the **Service**.

- ZooKeeper
- HDFS

Step 14 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 15 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.98 ALM-14020 Number of Entries in the HDFS Directory Exceeds the Threshold

Description

The system obtains the number of subfiles and subdirectories in a specified directory every hour and checks whether it reaches the percentage of the threshold (the maximum number of subfiles and subdirectories in an HDFS directory, the threshold for triggering an alarm is **90%** by default). If it exceeds the percentage of the threshold, an alarm is triggered.

When the number of subfiles and subdirectories in the directory the alarm is lower than the percentage of the threshold, the alarm is automatically cleared. When the monitoring switch is disabled, alarms corresponding to all directories are cleared. If a directory is removed from the monitoring list, alarms corresponding to the directory are cleared.

NOTE

- The **dfs.namenode.fs-limits.max-directory-items** parameter specifies the maximum number of subfiles and subdirectories in the HDFS directory. Its default value is **1048576**. If the number of subfiles and subdirectories in a directory exceeds the parameter value, subfiles and subdirectories cannot be created in the directory.
- The **dfs.namenode.directory-items.monitor** parameter specifies the list of directories to be monitored. Its default value is **/tmp,/SparkJobHistory,/mr-history**.
- The **dfs.namenode.directory-items.monitor.enabled** parameter is used to enable or disable the monitoring switch. Its default value is **true**, which means the monitoring switch is enabled by default.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14020	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
NameServiceName	Specifies the NameService service for which the alarm is generated.
Directory	Specifies the directory for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the number of entries in the monitored directory exceeds 90% of the threshold, an alarm is triggered, but entries can be added to the directory. Once the maximum threshold is exceeded, entries will fail to be added to the directory.

Possible Causes

The number of entries in the monitored directory exceeds 90% of the threshold.

Procedure

Check whether unnecessary files exist in the system.

Step 1 Log in to the HDFS client as user **root**. Run the **cd** command to go to the client installation directory, and run the **source bigdata_env** command to set the environment variables.

If the cluster is in security mode, security authentication is required.

Run the **kinit hdfs** command and enter the password as prompted. Obtain the password from the administrator.

Step 2 Run the following command to check whether files and directories in the directory with the alarm can be deleted:

hdfs dfs -ls *Directory with the alarm*

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 Run the following command to delete unnecessary files.

hdfs dfs -rm -r -f *File or directory path*

NOTE

Deleting a file or folder is a high-risk operation. Ensure that the file or folder is no longer required before performing this operation.

Step 4 Wait 1 hour and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check whether the threshold is correctly configured.

Step 5 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Configurations** > **All Configurations**. Search for the **dfs.namenode.fs-limits.max-directory-items** parameter and check whether the parameter value is appropriate.

- If yes, go to **Step 9**.
- If no, go to **Step 6**.

Step 6 Increase the parameter value.

Step 7 Save the configuration and click **Dashboard** > **More** > **Restart Service**.

Step 8 Wait 1 hour and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect fault information.

Step 9 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 10 Select **HDFS** in the required cluster from the **Service**.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.99 ALM-14021 NameNode Average RPC Processing Time Exceeds the Threshold

Description

The system checks the average RPC processing time of NameNode every 30 seconds, and compares the actual average RPC processing time with the threshold (default value: 100 ms). This alarm is generated when the system detects that the average RPC processing time exceeds the threshold for several consecutive times (10 times by default).

You can choose **O&M > Alarm > Thresholds >** *Name of the desired cluster* > **HDFS** to change the threshold.

When the **Trigger Count** is 1, this alarm is cleared when the average RPC processing time of NameNode is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the average RPC processing time of NameNode is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14021	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
NameServiceName	Specifies the NameService service for which the alarm is generated.
Trigger condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

NameNode cannot process the RPC requests from HDFS clients, upper-layer services that depend on HDFS, and DataNode in a timely manner. Specifically, the services that access HDFS run slowly or the HDFS service is unavailable.

Possible Causes

- The CPU performance of NameNode nodes is insufficient and therefore NameNode nodes cannot process messages in a timely manner.
- The configured NameNode memory is too small and frame freezing occurs on the JVM due to frequent full garbage collection.
- NameNode parameters are not configured properly, so NameNode cannot make full use of system performance.

Procedure

Obtain alarm information.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms**. In the alarm list, click the alarm.

Step 2 Check the alarm. Obtain the host name of the NameNode node involved in this alarm from the **HostName** information of **Location**. Then obtain the name of the NameService node involved in this alarm from the **NameServiceName** information of **Location**.

Check whether the threshold is too small.

Step 3 Check the status of the services that depend on HDFS. Check whether the services run slowly or task execution times out.

- If yes, go to **Step 8**.
- If no, go to **Step 4**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **RPC**, and select **Average Time of Active NameNode RPC Processing** and click **OK**.

Step 5 On the **Average Time of Active NameNode RPC Processing** monitoring page, obtain the value of the NameService node involved in this alarm.

Step 6 On the FusionInsight Manager portal, choose **O&M > Alarm > Thresholds >** *Name of the desired cluster* **> HDFS**. Locate **Average Time of Active NameNode RPC Processing** and click the **Modify** in the **Operation** column of the default rule. The **Modify Rule** page is displayed. Change **Threshold** to 150% of the peak value within one day before and after the alarm is generated. Click **OK** to save the new threshold.

Step 7 Wait for 5 minutes and then check whether the alarm is automatically cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Check whether the CPU performance of the NameNode node is sufficient.

Step 8 On the FusionInsight Manager portal, click **O&M > Alarm >Alarms** and check whether **ALM-12016 CPU Usage Exceeds the Threshold** is generated for the NameNode node.

- If yes, go to **Step 9**.
- If no, go to **Step 11**.

Step 9 Handle **ALM-12016 CPU Usage Exceeds the Threshold** by taking recommended actions.

Step 10 Wait for 10 minutes and check whether alarm 14021 is automatically cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Check whether the memory of the NameNode node is too small.

Step 11 On the FusionInsight Manager portal, click **O&M > Alarm >Alarms** and check whether **ALM-14007 HDFS NameNode Heap Memory Usage Exceeds the Threshold** is generated for the NameNode node.

- If yes, go to **Step 12**.
- If no, go to **Step 14**.

Step 12 Handle **ALM-14007 HDFS NameNode Heap Memory Usage Exceeds the Threshold** by taking recommended actions.

Step 13 Wait for 10 minutes and check whether alarm 14021 is automatically cleared.

- If yes, no further action is required.
- If no, go to **Step 14**.

Check whether NameNode parameters are configured properly.

Step 14 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Configurations** > **All Configurations**. Search for parameter **dfs.namenode.handler.count** and view its value. If the value is less than or equal to 128, change it to **128**. If the value is greater than 128 but less than 192, change it to **192**.

Step 15 Search for parameter **ipc.server.read.threadpool.size** and view its value. If the value is less than 5, change it to **5**.

Step 16 Click **Save** and click **OK**.

Step 17 On the **Instance** page of HDFS, select the standby NameNode of NameService involved in this alarm and choose **More** > **Restart Instance**. Enter the password and click **OK**. Wait until the standby NameNode is started up.

Step 18 On the **Instance** page of HDFS, select the active NameNode of NameService involved in this alarm and choose **More** > **Restart Instance**. Enter the password and click **OK**. Wait until the active NameNode is started up.

Step 19 Wait for 1 hour and then check whether the alarm is automatically cleared.

- If yes, no further action is required.
- If no, go to **Step 20**.

Collect fault information.

Step 20 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 21 Select the following node in the required cluster from the **Service**.

- HDFS

Step 22 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 23 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.100 ALM-14022 NameNode Average RPC Queuing Time Exceeds the Threshold

Description

The system checks the average RPC queuing time of NameNode every 30 seconds, and compares the actual average RPC queuing time with the threshold (default value: 200 ms). This alarm is generated when the system detects that the average RPC queuing time exceeds the threshold for several consecutive times (10 times by default).

You can choose **O&M > Alarm > Thresholds >** *Name of the desired cluster* > **HDFS** to change the threshold.

When the **Trigger Count** is 1, this alarm is cleared when the average RPC queuing time of NameNode is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the average RPC queuing time of NameNode is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14022	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
NameServiceName	Specifies the NameService service for which the alarm is generated.
Trigger condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

NameNode cannot process the RPC requests from HDFS clients, upper-layer services that depend on HDFS, and DataNode in a timely manner. Specifically, the services that access HDFS run slowly or the HDFS service is unavailable.

Possible Causes

- The CPU performance of NameNode nodes is insufficient and therefore NameNode nodes cannot process messages in a timely manner.
- The configured NameNode memory is too small and frame freezing occurs on the JVM due to frequent full garbage collection.
- NameNode parameters are not configured properly, so NameNode cannot make full use of system performance.
- The volume of services that access HDFS is too large and therefore NameNode is overloaded.

Procedure

Obtain alarm information.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms**. In the alarm list, click the alarm.

Step 2 Check the alarm. Obtain the alarm generation time from **Generated**. Obtain the host name of the NameNode node involved in this alarm from the **HostName** information of **Location**. Then obtain the name of the NameService node involved in this alarm from the **NameServiceName** information of **Location**.

Check whether the threshold is too small.

Step 3 Check the status of the services that depend on HDFS. Check whether the services run slowly or task execution times out.

- If yes, go to **Step 8**.
- If no, go to **Step 4**.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **RPC**, and select **Average Time of Active NameNode RPC Queuing** and click **OK**.

Step 5 On the **Average Time of Active NameNode RPC Queuing** monitoring page, obtain the value of the NameService node involved in this alarm.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **HDFS**. Locate **Average Time of Active NameNode RPC Queuing** and click the **Modify** in the **Operation** column of the default rule. The **Modify Rule** page is displayed. Change **Threshold** to 150% of the monitored value. Click **OK** to save the new threshold.

Step 7 Wait for 1 minute and then check whether the alarm is automatically cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Check whether the CPU performance of the NameNode node is sufficient.

Step 8 On the FusionInsight Manager portal, click **O&M** > **Alarm** > **Alarms** and check whether **ALM-12016 HDFS NameNode Memory Usage Exceeds the Threshold** is generated.

- If yes, go to **Step 9**.
- If no, go to **Step 11**.

Step 9 Handle **ALM-12016 CPU Usage Exceeds the Threshold** by taking recommended actions.

Step 10 Wait for 10 minutes and check whether alarm 14022 is automatically cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Check whether the memory of the NameNode node is too small.

Step 11 On the FusionInsight Manager portal, click **O&M > Alarm > Alarms** and check whether **ALM-14007 HDFS NameNode Memory Usage Exceeds the Threshold** is generated.

- If yes, go to **Step 12**.
- If no, go to **Step 14**.

Step 12 Handle **ALM-14007 CPU Usage Exceeds the Threshold** by taking recommended actions.

Step 13 Wait for 10 minutes and check whether alarm 14022 is automatically cleared.

- If yes, no further action is required.
- If no, go to **Step 14**.

Check whether NameNode parameters are configured properly.

Step 14 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS** > **Configurations** > **All Configurations**. Search for parameter **dfs.namenode.handler.count** and view its value. If the value is less than or equal to 128, change it to **128**. If the value is greater than 128 but less than 192, change it to **192**.

Step 15 Search for parameter **ipc.server.read.threadpool.size** and view its value. If the value is less than 5, change it to **5**.

Step 16 Click **Save**, and click **OK**.

Step 17 On the **Instance** page of HDFS, select the standby NameNode of NameService involved in this alarm and choose **More** > **Restart Instance**. Enter the password and click **OK**. Wait until the standby NameNode is started up.

Step 18 On the **Instance** page of HDFS, select the active NameNode of NameService involved in this alarm and choose **More** > **Restart Instance**. Enter the password and click **OK**. Wait until the active NameNode is started up.

Step 19 Wait for 1 hour and then check whether the alarm is automatically cleared.

- If yes, no further action is required.
- If no, go to **Step 20**.

Check whether the HDFS workload changes and reduce the workload properly.

Step 20 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **HDFS**. Click the drop-down menu in the upper right corner of **Chart**, click **Customize**, select **Average Time of Active NameNode RPC Queuing** and click **OK**.

Step 21 Click ⤡. The **Details** page is displayed.

Step 22 Set the monitoring data display period, from 5 days before the alarm generation time to the alarm generation time. Click **OK**.

Step 23 On the **Average RPC Queuing Time** monitoring page, check whether the point in time when the queuing time increases abruptly exists.

- If yes, go to **Step 24**.
- If no, go to **Step 27**.

Step 24 Confirm and check the point in time. Check whether a new task frequently accesses HDFS and whether the access frequency can be reduced.

Step 25 If a Balancer task starts at the point in time, stop the task or specify a node for the task to reduce the HDFS workload.

Step 26 Wait for 1 hour and then check whether the alarm is automatically cleared.

- If yes, no further action is required.
- If no, go to **Step 27**.

Collect fault information.

Step 27 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 28 Select **HDFS** in the required cluster from the **Service**.

Step 29 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 30 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.101 ALM-14023 Percentage of Total Reserved Disk Space for Replicas Exceeds the Threshold

Description

The system checks the percentage of total reserved disk space for replicas (Total reserved disk space for replicas/(Total reserved disk space for replicas + Total remaining disk space)) every 30 seconds and compares the actual percentage with the threshold (**90%** by default). This alarm is generated when the percentage of total reserved disk space for replicas exceeds the threshold for multiple consecutive times (**Trigger Count**).

The alarm is cleared in the following two scenarios: The value of **Trigger Count** is **1** and the percentage of total reserved disk space for replicas is less than or equal to the threshold; the value of **Trigger Count** is greater than **1** and the percentage of total reserved disk space for replicas is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14023	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
NameServiceName	Specifies the NameService service for which the alarm is generated.
Trigger condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

The performance of writing data to HDFS is affected. If all remaining DataNode space is reserved for replicas, writing HDFS data fails.

Possible Causes

- The alarm threshold is improperly configured.
- The disk space configured for the HDFS cluster is insufficient.
- The volume of services that access HDFS is too large and therefore DataNode is overloaded.

Procedure

Check whether the alarm threshold is appropriate.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **HDFS** > **Disk** > **Percentage of Reserved Space for Replicas of Unused Space** to check whether the alarm threshold is appropriate. (The default threshold is **90%**. Users can change it as required.)

- If yes, go to **Step 4**.
- If no, go to **Step 2**.

Step 2 Choose **O&M > Alarm > Thresholds >** *Name of the desired cluster* > **HDFS** > **Disk** > **Percentage of Reserved Space for Replicas of Unused Space** and Click **Modify,** change the threshold based on the actual usage.

Step 3 Wait 5 minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check whether an alarm indicating insufficient disk space is generated.

Step 4 On the FusionInsight Manager portal, check whether **ALM-14001 HDFS Disk Usage Exceeds the Threshold** or **ALM-14002 DataNode Disk Usage Exceeds the Threshold** exists on the **O&M > Alarm > Alarms** page.

- If yes, go to **Step 5**.
- If no, go to **Step 7**.

Step 5 Handle the alarm by referring to instructions in **ALM-14001 HDFS Disk Usage Exceeds the Threshold** or **ALM-14002 DataNode Disk Usage Exceeds the Threshold** and check whether the alarm is cleared.

- If yes, go to **Step 6**.
- If no, go to **Step 7**.

Step 6 Wait 5 minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Expand the DataNode capacity.

Step 7 Expand the DataNode capacity.

Step 8 Wait 5 minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect fault information.

Step 9 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 10 Select **HDFS** in the required cluster from the **Service**.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 20 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.102 ALM-14024 Tenant Space Usage Exceeds the Threshold

Description

The system checks the space usage (used space of each directory/space allocated to each directory) of each directory associated with a tenant every hour and compares the space usage of each directory with the threshold set for the directory. This alarm is generated when the space usage exceeds the threshold.

This alarm is cleared when the space usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14024	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
TenantName	Specifies the tenant for which the alarm is generated.
DirectoryName	Specifies the directory for which the alarm is generated.
Trigger condition	Specifies the threshold for triggering the alarm.

Impact on the System

This alarm is generated if the space usage of the tenant directory exceeds the custom threshold. File writing to the directory is not affected. If the used space exceeds the maximum storage space allocated to the directory, the HDFS fails to write data to the directory.

Possible Causes

- The alarm threshold is improperly configured.
- The space allocated to the tenant is improper.

Procedure

Check whether the alarm threshold is appropriate.

Step 1 View the alarm location information to obtain the tenant name and tenant directory for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose the **Tenant Resources** page, select the tenant for which the alarm is generated, and click **Resources**. Check whether the storage space threshold configured for the tenant directory for which the alarm is generated is proper. (The default value 90% is a proper value. You can set it based on the site requirements.)

- If yes, go to **Step 5**.
- If no, go to **Step 3**.

Step 3 On the **Resources** page, click **Modify** to modify or delete the storage space threshold.

Step 4 About one minute later, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check whether the space allocated to the tenant is appropriate.

Step 5 On the FusionInsight Manager portal, choose the **Tenant Resources** page, select the tenant for which the alarm is generated, and click **Resources**. Check whether the storage space quota of the tenant directory for which the alarm is generated is proper based on the actual service status of the tenant directory.

- If yes, go to **Step 8**.
- If no, go to **Step 6**.

Step 6 On the **Resources** page, click **Modify** to modify the storage space quota.

Step 7 About one minute later, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Collect fault information.

Step 8 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 9 Select **HDFS** in the required cluster and **NodeAgent** under **Manager** from the **Service**.

Step 10 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 20 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.103 ALM-14025 Tenant File Object Usage Exceeds the Threshold

Description

The system checks the file object usage (used file objects of each directory/number of file objects allocated to each directory) of each directory associated with a tenant every hour and compares the file object usage of each directory with the threshold set for the directory. This alarm is generated when the file object usage exceeds the threshold.

This alarm is cleared when the file object usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
14025	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
TenantName	Specifies the tenant for which the alarm is generated.
DirectoryName	Specifies the directory for which the alarm is generated.
Trigger condition	Specifies the threshold for triggering the alarm.

Impact on the System

This alarm is generated if the usage of file objects in a tenant directory exceeds the custom threshold. File writing to the directory is not affected. If the number of used file objects exceeds the maximum number of file objects allocated to the directory, the HDFS fails to write data to the directory.

Possible Causes

- The alarm threshold is improperly configured.
- The maximum number of file objects allocated to the tenant directory is inappropriate.

Procedure

Check whether the alarm threshold is appropriate.

Step 1 View the alarm location information to obtain the tenant name and tenant directory for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose the **Tenant Resources** page, select the tenant for which the alarm is generated, and click **Resources**. Check whether the file object threshold configured for the tenant directory for which the alarm is generated is proper. (The default value 90% is a proper value. You can set it based on the site requirements.)

- If yes, go to **Step 5**.
- If no, go to **Step 3**.

Step 3 On the **Resources** page, click **Modify** to modify or delete the file object threshold of the tenant directory for which the alarm is generated.

Step 4 About one minute later, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check whether the maximum number of file objects allocated to the tenant is appropriate.

Step 5 On the FusionInsight Manager portal, choose the **Tenant Resources** page, select the tenant for which the alarm is generated, and click **Resources**. Check whether the maximum number of file objects configured for the tenant directory for which the alarm is generated is proper based on the actual service status of the tenant directory.

- If yes, go to **Step 8**.
- If no, go to **Step 6**.

Step 6 On the **Resources** page, click **Modify** to modify or delete the maximum number of file objects configured for the tenant directory.

Step 7 About one minute later, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Collect fault information.

Step 8 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 9 Select **HDFS** in the required cluster and **NodeAgent** under **Manager** from the **Service**.

Step 10 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 20 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.104 ALM-14026 Blocks on DataNode Exceed the Threshold

Description

The system checks the number of blocks on each DataNode every 30 seconds. This alarm is generated when the number of blocks on the DataNode exceeds the threshold.

If **Trigger Count** is **1** and the number of blocks on the DataNode is less than or equal to the threshold, this alarm is cleared. If **Trigger Count** is greater than **1** and the number of blocks on the DataNode is less than or equal to 90% of the threshold, this alarm is cleared.

Attribute

Alarm ID	Alarm Severity	Auto Clear
14026	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.

Name	Meaning
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger condition	Specifies the threshold for triggering the alarm.

Impact on the System

If this alarm is reported, there are too many blocks on the DataNode. In this case, data writing into the HDFS may fail due to insufficient disk space.

Possible Causes

- The alarm threshold is improperly configured.
- Data skew occurs among DataNodes.
- The disk space configured for the HDFS cluster is insufficient.

Procedure

Change the threshold.

Step 1 On FusionInsight Manager, choose **Cluster**, click the name of the desired cluster, and choose **HDFS**. Then choose **Configurations** > **All Configurations**. On the displayed page, find the **GC_OPTS** parameter under **HDFS->DataNode**.

Step 2 Set the threshold of the DataNode blocks. Specifically, change the value of **Xmx** of the **GC_OPTS** parameter. **Xmx** specifies the memory, and each GB memory supports a maximum of 500,000 DataNode blocks. Set the memory as required. Confirm that **GC_PROFILE** is set to **custom** and save the configuration.

Step 3 Choose **Cluster**, click the name of the desired cluster, and choose **HDFS** > **Instance**. Select the DataNode instance whose status is **Expired**, click **More**, and select **Restart Instance** to make the **GC_OPTS** configuration take effect.

Step 4 Check whether the alarm is cleared 5 minutes later.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check whether associated alarms are reported.

Step 5 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**, and check whether the **ALM-14002 DataNode Disk Usage Exceeds the Threshold** alarm exists.

- If yes, go to **Step 6**.
- If no, go to **Step 8**.

Step 6 Handle the alarm by following the instructions in **ALM-14002 DataNode Disk Usage Exceeds the Threshold** and check whether the alarm is cleared.

- If yes, go to **Step 7**.
- If no, go to **Step 8**.

Step 7 Check whether the alarm is cleared 5 minutes later.

- If yes, no further action is required.
- If no, go to **Step 8**.

Expand the DataNode capacity.

Step 8 Expand the DataNode capacity.

Step 9 On FusionInsight Manager, wait for 5 minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 10**.

Collect the fault information.

Step 10 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 11 Expand the drop-down list next to the **Service** field. In the **Services** dialog box that is displayed, select **HDFS** for the target cluster.

Step 12 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 20 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 13 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

Configuration rules of the DataNode JVM parameter.

Default value of the DataNode JVM parameter **GC_OPTS**:

-Xms2G -Xmx4G -XX:NewSize=128M -XX:MaxNewSize=256M -XX:MetaspaceSize=128M -XX:MaxMetaspaceSize=128M -XX:+UseConcMarkSweepGC -XX:+CMSParallelRemarkEnabled -XX:CMSInitiatingOccupancyFraction=65 -XX:+PrintGCDetails -Dsun.rmi.dgc.client.gcInterval=0x7FFFFFFFFFFFFFE -Dsun.rmi.dgc.server.gcInterval=0x7FFFFFFFFFFFFFE -XX:-OmitStackTraceInFastThrow -XX:+PrintGCDateStamps -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=1M -Djdk.tls.ephemeralDHKeySize=2048

The average number of blocks stored in each DataNode instance in the cluster is: Number of HDFS blocks x 3/Number of DataNodes. If the average number of blocks changes, you need to change **-Xms2G -Xmx4G -XX:NewSize=128M -**

XX:MaxNewSize=256M in the default value. The following table lists the reference values.

Table 9-91 DataNode JVM configuration

Average Number of Blocks in a DataNode Instance	Reference Value
2,000,000	-Xms6G -Xmx6G -XX:NewSize=512M -XX:MaxNewSize=512M
5,000,000	-Xms12G -Xmx12G -XX:NewSize=1G -XX:MaxNewSize=1G

Xmx specifies memory which corresponds to the threshold of the number of DataNode blocks, and each GB memory supports a maximum of 500,000 DataNode blocks. Set the memory as required.

9.13.105 ALM-14027 DataNode Disk Fault

Description

The system checks the disk status on DataNodes every 60 seconds. This alarm is generated when a disk is faulty.

After all faulty disks on the DataNode are recovered, you need to manually clear the alarm and restart the DataNode.

Attribute

Alarm ID	Alarm Severity	Auto Clear
14027	Major	No

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Name	Meaning
Failed Volumes	Specifies the list of faulty disks.

Impact on the System

If this alarm is reported, there are abnormal disk partitions on the DataNode. This may cause the loss of written files.

Possible Causes

- The hard disk is faulty.
- The disk permissions are configured improperly.

Procedure

Check whether a disk alarm is generated.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** and check whether **ALM-12014 Partition Lost** or **ALM-12033 Slow Disk Fault** exists.

- If yes, go to **Step 2**.
- If no, go to **Step 4**.

Step 2 Rectify the fault by referring to the handling procedure of **ALM-12014 Partition Lost** or **ALM-12033 Slow Disk Fault**. Then, check whether the alarm is cleared.

- If yes, go to **Step 3**.
- If no, go to **Step 4**.

Step 3 Wait 5 minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Modify disk permissions.

Step 4 Choose **O&M** > **Alarm** > **Alarms** and view **Location** and **Additional Information** of the alarm to obtain the location of the faulty disk.

Step 5 Log in to the node for which the alarm is generated as user **root**. Go to the directory where the faulty disk is located, and run the **ll** command to check whether the permission of the faulty disk is **711** and whether the user is **omm**.

- If yes, go to **Step 8**.
- If no, go to **Step 6**.

Step 6 Modify the permission of the faulty disk. For example, if the faulty disk is **data1**, run the following commands:

```
chown omm:wheel data1
```

```
chmod 711 data1
```

Step 7 In the alarm list on Manager, click **Clear** in the **Operation** column of the alarm to manually clear the alarm. Choose **Cluster** > **Services** > **HDFS** > **Instance**, select the DataNode, choose **More** > **Restart Instance**, wait for 5 minutes, and check whether a new alarm is reported.

- If no, no further action is required.
- If yes, go to **Step 8**.

Collect the fault information.

Step 8 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 9 Expand the **Service** drop-down list, and select **HDFS** and **OMS** for the target cluster.

Step 10 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 20 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system does not automatically clear this alarm and you need to manually clear the alarm.

Related Information

None

9.13.106 ALM-14028 Number of Blocks to Be Supplemented Exceeds the Threshold

Description

The system checks the number of blocks to be supplemented every 30 seconds and compares the number with the threshold. The number of blocks to be supplemented has a default threshold. This alarm is generated when the number of blocks to be supplemented exceeds the threshold.

You can change the threshold specified by **Blocks Under Replicated (NameNode)** by choosing **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **HDFS** > **File and Block**.

If **Trigger Count** is set to **1** and the number of blocks to be supplemented is less than or equal to the threshold, this alarm is cleared. If **Trigger Count** is greater than **1** and the number of blocks to be supplemented is less than or equal to 90% of the threshold, this alarm is cleared.

Attribute

Alarm ID	Alarm Severity	Auto Clear
14028	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
NameServiceName	Specifies the NameService for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Data stored in HDFS is lost. HDFS may enter the security mode and cannot provide write services. Lost block data cannot be restored.

Possible Causes

- The DataNode instance is abnormal.
- Data is deleted.
- The number of replicas written into the file is greater than the number of DataNodes.

Procedure

Step 1 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Alarm** > **Alarms**. On the page that is displayed, check whether alarm **ALM-14003 Number of Lost HDFS Blocks Exceeds the Threshold** is generated.

- If yes, go to **Step 2**.
- If no, go to **Step 3**.

Step 2 Rectify the fault according to the handling procedure of **ALM-14003 Number of Lost HDFS Blocks Exceeds the Threshold**. Five minutes later, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 3**.

Step 3 Log in to the HDFS client as user **root**. The user password is defined by the user before the installation. Contact the MRS cluster administrator to obtain the password. Run the following commands:

- Security mode:

  cd *Client installation directory*

  source bigdata_env

  kinit hdfs

- Normal mode:

  su - omm

  cd *Client installation directory*

  source bigdata_env

Step 4 Run the **hdfs fsck / >> fsck.log** command to obtain the status of the current cluster.

Step 5 Run the following command to count the number (*M*) of blocks to be replicated:

cat fsck.log | grep "Under-replicated"

Step 6 Run the following command to count the number (*N*) of blocks to be replicated in the **/tmp/hadoop-yarn/staging/** directory:

cat fsck.log | grep "Under replicated" | grep "/tmp/hadoop-yarn/staging/" | wc -l

NOTE

/tmp/hadoop-yarn/staging/ is the default directory. If the directory is modified, obtain it from the configuration item **yarn.app.mapreduce.am.staging-dir** in the **mapred-site.xml** file.

Step 7 Check whether the percentage of *N* is greater than 50% (N/M > 50%).

- If yes, go to **Step 8**.
- If no, go to **Step 9**.

Step 8 Run the following command to reconfigure the number of file replicas in the directory (set the number of file replicas to the number of DataNodes or the default number of file replicas):

hdfs dfs -setrep -w *Number of file replicas*/tmp/hadoop-yarn/staging/

NOTE

To obtain the default number of file replicas:

Log in to FusionInsight Manager, choose **Cluster** > **Services** > **HDFS** > **Configurations** > **All Configurations**, and search for the **dfs.replication** parameter. The value of this parameter is the default number of file replicas.

Check whether the alarm is cleared 5 minutes later.

- If yes, no further action is required.

- If no, go to **Step 9**.

Collect the fault information.

Step 9 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 10 Expand the drop-down list next to the **Service** field. In the **Services** dialog box that is displayed, select **HDFS** for the target cluster.

Step 11 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.107 ALM-14029 Number of Blocks in a Replica Exceeds the Threshold

Description

The system checks the number of blocks in a single replica every four hours and compares the number with the threshold. There is a threshold for the number of blocks in a single replica. This alarm is generated when the actual number of blocks in a single replica exceeds the threshold.

This alarm is cleared when the number of blocks to be supplemented is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
14029	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
NameServiceName	Specifies the NameService for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Replica data is prone to be lost when a node is faulty. Too many files of a single replica affect the security of the HDFS file system.

Possible Causes

- The DataNode is faulty.
- The disk is faulty.
- Files are written to a single replica.

Procedure

Step 1 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Alarm** > **Alarms**. On the page that is displayed, check whether alarm **ALM-14003 Number of Lost HDFS Blocks Exceeds the Threshold** is generated.

- If yes, go to **Step 2**.
- If no, go to **Step 3**.

Step 2 Rectify the fault according to the handling procedure of **ALM-14003 Number of Lost HDFS Blocks Exceeds the Threshold**. In the next detection period, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 3**.

Step 3 Check whether files of a single replica have been written into the service.

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 4 Log in to the HDFS client as user **root**. The user password is defined by the user before the installation. Contact the MRS cluster administrator to obtain the password. Run the following commands:

- Security mode:

  cd *Client installation directory*

  source bigdata_env

  kinit hdfs

- Normal mode:

  su - omm

  cd *Client installation directory*

  source bigdata_env

Step 5 Run the following command on the client node to increase the number of replicas for a single replica file:

hdfs dfs -setrep -w *file replica number file name or file path*

Step 6 In the next detection period, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect the fault information.

Step 7 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 8 Expand the drop-down list next to the **Service** field. In the **Services** dialog box that is displayed, select **HDFS** for the target cluster.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.108 ALM-14030 HDFS Allows Write of Single-Replica Data

Description

This alarm is generated when **dfs.single.replication.enable** is set to **true**, indicating that HDFS is configured to allow write of single-replica data.

This alarm is cleared when this function is disabled on HDFS.

Attribute

Alarm ID	Alarm Severity	Auto Clear
14030	Warning	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.

Impact on the System

Data of a single replica may be lost. Therefore, the system does not allow write of single-replica data by default. If this configuration is enabled on HDFS and the number of HDFS replicas configured on the client is 1, single-replica data can be written to HDFS.

Possible Causes

The HDFS configuration item **dfs.single.replication.enable** is set to **true**.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **HDFS**. On the page that is displayed, click the **Configurations** tab then the **All Configurations** sub-tab.

Step 2 Search for **dfs.single.replication.enable** in the search box, change the value of the configuration item to **false**, and click **Save**.

Step 3 On the **Dashboard** page of the HDFS service, click **More** and select **Service Rolling Restart** in the upper right corner.

Step 4 After the HDFS service is started, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Collect fault information.

Step 5 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 6 Expand the drop-down list next to the **Service** field. In the **Services** dialog box that is displayed, select **HDFS** for the target cluster.

Step 7 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 8 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.109 ALM-16000 Percentage of Sessions Connected to the HiveServer to Maximum Number Allowed Exceeds the Threshold

Description

The system detects the percentage of sessions connected to the HiveServer to the maximum number of allowed sessions every 30 seconds. This indicator can be viewed on the **Cluster** > *Name of the desired cluster* > **Services** > **Hive > Instance** > *HiveServer instance***.** This alarm is generated when the percentage exceeds the default value **90%**.

To change the threshold, choose **O&M > Alarm > Thresholds >** *Name of the desired cluster* > **Hive > Percentage of Sessions Connected to the HiveServer to Maximum Number of Sessions Allowed by the HiveServer**.

When the **Trigger Count** is 1, this alarm is cleared when the percentage is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the percentage is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
16000	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.

Name	Meaning
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If a connection alarm is generated, too many sessions are connected to Hive and new connections are unavailable.

Possible Causes

Too many clients are connected to HiveServer.

Procedure

Increase the maximum number of connections to Hive.

Step 1 On the FusionInsight Manager portal, Choose **Cluster** > *Name of the desired cluster* > **Services** > **Hive** > **Configurations >All Configurations**.

Step 2 Search for **hive.server.session.control.maxconnections** and increase the value of this parameter. If the value of this parameter is **A**, the threshold is **B**, and the number of sessions connected to the HiveServer is **C**, adjust the value of this parameter according to **A x B > C**. To view the number of sessions connected to the HiveServer, check the value of **Statistics for Sessions of the HiveServer** on the Hive monitoring page.

Step 3 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Collect fault information.

Step 4 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 5 Select **Hive** in the required cluster from the **Service**.

Step 6 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 7 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.110 ALM-16001 Hive Warehouse Space Usage Exceeds the Threshold

Description

This alarm is generated when the Hive warehouse space usage exceeds the specified threshold (85% by default). The system checks the Hive data warehouse space usage every 30s. The indicator **Percentage of HDFS Space Used by Hive to the Available Space** can be viewed on the Hive service monitoring page.

To change the threshold, choose **O&M > Alarm > Thresholds >** *Name of the desired cluster* > **Hive > Percentage of HDFS Space Used by Hive to the Available Space**.

When the **Trigger Count** is 1, this alarm is cleared when the Hive warehouse space usage is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the Hive warehouse space usage is less than or equal to 90% of the threshold.

NOTE

The administrator can reduce the warehouse space usage by expanding the warehouse capacity or releasing the used space.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
16001	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Name	Meaning
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

The system fails to write data, which causes data loss.

Possible Causes

- The upper limit of the HDFS capacity available for Hive is too small.
- The HDFS space is insufficient.
- Some data nodes break down.

Procedure

Expand the system configuration.

Step 1 Analyze the cluster HDFS capacity usage and increase the upper limit of the HDFS capacity available for Hive.

Log in to FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Hive** > **Configurations > All Configurations**, find **hive.metastore.warehouse.size.percent**, and increase its value so that larger HDFS capacity will be available for Hive. Assume that the value of the configuration item is A, the total HDFS storage space is B, the threshold is C, and the HDFS space used by Hive is D. The adjustment policy is A x B x C > D. The total HDFS storage space can be viewed on the HDFS NameNode page. The HDFS space used by Hive can be viewed on the Hive monitoring page.

Step 2 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 3**.

Expand the system.

Step 3 Expand the system.

Step 4 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check whether the data node is normal.

Step 5 On the FusionInsight Manager portal, click **O&M > Alarm > Alarms**.

Step 6 Check whether "ALM-12006 Node Fault", "ALM-12007 Process Fault", or "ALM-14002 DataNode Disk Usage Exceeds the Threshold" exist.

- If yes, go to **Step 7**.

- If no, go to **Step 9**.

Step 7 Clear the alarm by following the steps provided in "ALM-12006 Node Fault", "ALM-12007 Process Fault", and "ALM-14002 DataNode Disk Usage Exceeds the Threshold".

Step 8 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect fault information.

Step 9 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 10 Select **Hive** in the required cluster from the **Service**.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.111 ALM-16002 Hive SQL Execution Success Rate Is Lower Than the Threshold

Description

The system checks the percentage of the HQL statements that are executed successfully in every 30 seconds. The formula is: Percentage of HQL statements that are executed successfully = Number of HQL statements that are executed successfully by Hive in a specified period/Total number of HQL statements that are executed by Hive. This indicator can be viewed on the **Cluster** > *Name of the desired cluster* > **Services** > **Hive** > **Instance** > *HiveServer instance* **.** The default threshold of the percentage of HQL statements that are executed successfully is **90%**. An alarm is reported when the percentage is lower than the **90%**. Users can view the name of the host where an alarm is generated in the location information about the alarm. The IP address of the host is the IP address of the HiveServer node.

Users can modify the threshold by choosing **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Hive** > **Percentage of HQL Statements That Are Executed Successfully by Hive**.

This alarm is cleared when the execution success rate is higher than 110% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
16002	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

The system configuration and performance cannot meet service processing requirements.

Possible Causes

- A syntax error occurs in HQL statements.
- The HBase service is abnormal when a Hive on HBase task is performed.
- The Spark service is abnormal when a Hive on Spark task is performed.
- The dependent basic services, such as HDFS, Yarn, and ZooKeeper, are abnormal.

Procedure

Check whether the HQL statements comply with syntax.

Step 1 On the FusionInsight Manager page, choose **O&M** > **Alarm** to view the alarm details and obtain the node where the alarm is generated.

Step 2 Use the Hive client to log in to the HiveServer node where an alarm is reported. Query the HQL syntax provided by Apache, and check whether the HQL commands are correct.

- If yes, go to **Step 4**.
- If no, go to **Step 3**.

NOTE

To view the user who runs an incorrect statement, you can download the hiveserver audit log file of the HiveServer node where this alarm is generated. **Start Data** and **End Data** are 10 minutes before and after the alarm generation time respectively. Open the log file and search for the **Result=FAIL** keyword to filter the log information about the incorrect statement, and then view the user who runs the incorrect statement according to **UserName** in the log information.

Step 3 Enter the correct HQL statements, and check whether the command can be properly executed.

- If yes, go to **Step 12**.
- If no, go to **Step 4**.

Check whether the HBase service is abnormal.

Step 4 Check whether an Hive on HBase task is performed with the user who runs the HQL command.

- If yes, go to **Step 5**.
- If no, go to **Step 8**.

Step 5 On the FusionInsight Manager page, click **Cluster** > *Name of the desired cluster* > **Services**, check whether the HBase service is normal in the service list.

- If yes, go to **Step 8**.
- If no, go to **Step 6**.

Step 6 Choose **O&M** > **Alarm**, check the related alarms displayed on the alarm page and clear them according to related alarm help.

Step 7 Enter the correct HQL statements, and check whether the command can be properly executed.

- If yes, go to **Step 12**.
- If no, go to **Step 8**.

Check whether the HDFS, Yarn, and ZooKeeper are normal.

Step 8 On the FusionInsight Manager portal, click **Cluster** > *Name of the desired cluster* > **Services**.

Step 9 In the service list, check whether the services, such as HDFS, Yarn, and ZooKeeper are normal.

- If yes, go to **Step 12**.
- If no, go to **Step 10**.

Step 10 Check the related alarms displayed on the alarm page and clear them according to related alarm help.

Step 11 Enter the correct HQL statements, and check whether the command can be properly executed.

- If yes, go to **Step 12**.
- If no, go to **Step 13**.

Step 12 After 1 minute, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 13**.

Collect fault information.

Step 13 On the FusionInsight Manager home page, choose **O&M** > **Log > Download**.

Step 14 Select the following nodes in the required cluster from the **Service**:

- MapReduce
- Hive

Step 15 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 16 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.112 ALM-16003 Background Thread Usage Exceeds the Threshold

Description

The system checks the background thread usage in every 30 seconds. This alarm is generated when the usage of the background thread pool of Hive exceeds the threshold, 90% by default.

NOTE

MRS 3.X supports the multi-instance function. If the multi-instance function is enabled in the cluster and multiple Hive services are installed, determine the Hive service for which the alarm is generated based on the value of **ServiceName** in **Location** of the alarm. For example, if Hive1 service is unavailable, **ServiceName** is set to **Hive1** in **Location**, and the operation object in the handling procedure is changed from Hive to Hive1.

Attribute

Alarm ID	Alarm Severity	Auto Clear
16003	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

There are too many background threads, so the newly submitted task cannot run in time.

Possible Causes

The usage of the background thread pool of Hive is excessively high when:

- There are many tasks executed in the background thread pool of HiveServer.
- The capacity of the background thread pool of HiveServer is too small.

Procedure

Check the number of tasks executed in the background thread pool of HiveServer.

Step 1 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Hive**. On the displayed page, click **HiveServer Instance** and check values of **Background Thread Count** and **Background Thread Usage**.

Step 2 Check whether the number of background threads in the latest half an hour is excessively high. (By default, the queue number is 100, and the thread number is considered as high if it is 90 or larger.)

- If it is, go to **Step 3**.
- If it is not, go to **Step 5**.

Step 3 Adjust the number of tasks submitted to the background thread pool. (For example, cancel some time-consuming tasks with low performance.)

Step 4 Check whether the values of Background Thread Count and Background Thread Usage decrease.

- If it is, go to **Step 7**.
- If it is not, go to **Step 5**.

Check the capacity of the HiveServer background thread pool.

Step 5 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Hive**. On the displayed page, click **HiveServer Instance** and check values of Background Thread Count and Background Thread Usage.

Step 6 Increase the value of **hive.server2.async.exec.threads** in the **${BIGDATA_HOME}/FusionInsight_HD_8.1.0.1/1_23_HiveServer/etc/hive-site.xml** file. For example, increase the value by 20%.

Step 7 Save the modification.

Step 8 Check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 9**.

Collect fault information.

Step 9 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 10 Select **Hive** in the required cluster from the **Service**.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.113 ALM-16004 Hive Service Unavailable

Description

This alarm is generated when the HiveServer service is unavailable. The system checks the HiveServer service status every 60 seconds.

This alarm is cleared when the HiveServer service is normal.

NOTE

MRS 3.X supports the multi-instance function. If the multi-instance function is enabled in the cluster and multiple Hive service instances are installed, you need to determine the Hive service instance where the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the Hive1 service is unavailable, **ServiceName=Hive1** is displayed in **Location**, and the operation object in the procedure needs to be changed from Hive to Hive1.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
16004	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The system cannot provide data loading, query, and extraction services.

Possible Causes

- Hive service unavailability may be related to the faults of the Hive process as well as basic services, such as ZooKeeper, Hadoop distributed file system (HDFS), Yarn, and DBService.
  - The ZooKeeper service is abnormal.
  - The HDFS service is abnormal.
  - The Yarn service is abnormal.
  - The DBService service is abnormal.
  - The Hive service process is abnormal. If the alarm is caused by Hive process fault, the alarm report has a delay of about 5 minutes.
- The network communication between the Hive and basic services is interrupted.

Procedure

Check the HiveServer/MetaStore process status.

Step 1 On the FusionInsight Manager portal, click **Cluster >** *Name of the desired cluster* **> Services** > **Hive** > **Instance**. In the Hive instance list, check whether the HiveServer or MetaStore instances are in the Unknown state.

- If yes, go to **Step 2**.
- If no, go to **Step 4**.

Step 2 In the Hive instance list, choose **More** > **Restart Instance** to restart the HiveServer/MetaStore process.

Step 3 In the alarm list, check whether **Hive Service Unavailable** is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check the ZooKeeper service status.

Step 4 On the FusionInsight Manager, check whether the alarm list contains **Process Fault**.

- If yes, go to **Step 5**.
- If no, go to **Step 8**.

Step 5 In the **Process Fault**, check whether **ServiceName** is **ZooKeeper**.

- If yes, go to **Step 6**.
- If no, go to **Step 8**.

Step 6 Rectify the fault by following the steps provided in "ALM-12007 Process Fault".

Step 7 In the alarm list, check whether **Hive Service Unavailable** is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Check the HDFS service status.

Step 8 On the FusionInsight Manager, check whether the alarm list contains **HDFS Service Unavailable**.

- If yes, go to **Step 9**.
- If no, go to **Step 11**.

Step 9 Rectify the fault by following the steps provided in "ALM-14000 HDFS Service Unavailable".

Step 10 In the alarm list, check whether **Hive Service Unavailable** is cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Check the Yarn service status.

Step 11 In FusionInsight Manager alarm list, check whether **Yarn Service Unavailable** is generated.

- If yes, go to **Step 12**.

- If no, go to **Step 14**.

Step 12 Rectify the fault. For details, see "ALM-18000 Yarn Service Unavailable".

Step 13 In the alarm list, check whether **Hive Service Unavailable** is cleared.

- If yes, no further action is required.
- If no, go to **Step 14**.

Check the DBService service status.

Step 14 In FusionInsight Manager alarm list, check whether **DBService Service Unavailable** is generated.

- If yes, go to **Step 15**.
- If no, go to **Step 17**.

Step 15 Rectify the fault. For details, see "ALM-27001 DBService Service Unavailable".

Step 16 In the alarm list, check whether **Hive Service Unavailable** is cleared.

- If yes, no further action is required.
- If no, go to **Step 17**.

Check the network connection between the Hive and ZooKeeper, HDFS, Yarn, and DBService.

Step 17 On the FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Hive**.

Step 18 Click **Instance**.

The HiveServer instance list is displayed.

Step 19 Click **Host Name** in the row of **HiveServer**.

The active HiveServer host status page is displayed.

Step 20 Record the IP address under **Basic Information**.

Step 21 Use the IP address obtained in **Step 20** to log in to the host where the active HiveServer runs as user **omm**.

Step 22 Run the **ping** command to check whether communication between the host that runs the active HiveServer and the hosts that run the ZooKeeper, HDFS, Yarn, and DBService services is normal. (Obtain the IP addresses of the hosts that run the ZooKeeper, HDFS, Yarn, and DBService services in the same way as that for obtaining the IP address of the active HiveServer.)

- If yes, go to **Step 25**.
- If no, go to **Step 23**.

Step 23 Contact the administrator to restore the network.

Step 24 In the alarm list, check whether **Hive Service Unavailable** is cleared.

- If yes, no further action is required.
- If no, go to **Step 25**.

Collect fault information.

Step 25 On the FusionInsight Manager, choose **O&M** > **Log > Download**.

Step 26 Select the following nodes in the required cluster from the **Service**:

- ZooKeeper
- HDFS
- Yarn
- DBService
- Hive

Step 27 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 28 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.114 ALM-16005 The Heap Memory Usage of the Hive Process Exceeds the Threshold

Description

The system checks the Hive service status every 30 seconds. The alarm is generated when the heap memory usage of an Hive service exceeds the threshold (95% of the maximum memory).

Users can choose **O&M > Alarm > Thresholds >** *Name of the desired cluster* > **Hive** to change the threshold.

The alarm is cleared when the heap memory usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
16005	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.

Impact on the System

When the heap memory usage of Hive is overhigh, the performance of Hive task operation is affected. In addition, a memory overflow may occur so that the Hive service is unavailable.

Possible Causes

The heap memory of the Hive instance on the node is overused or the heap memory is inappropriately allocated. As a result, the usage exceeds the threshold.

Procedure

Check heap memory usage.

Step 1 On the FusionInsight Manager portal, click **O&M > Alarm > Alarms** and select the alarm whose **Alarm ID** is **16005**. Then check the role name in **Location** and confirm the IP adress of the instance.

- If the role for which the alarm is generated is HiveServer, go to **Step 2**.
- If the role for which the alarm is generated is MetaStore, go to **Step 3**.

Step 2 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* **> Services** > **Hive** > **Instance** and click the HiveServer for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **CPU and Memory**, and select **HiveServer Memory Usage Statistics** and click **OK**, check whether the used heap memory of the HiveServer service reaches the threshold(default value: 95%) of the maximum heap memory specified for HiveServer.

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 3 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Hive** > **Instance** and click the MetaStore for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **CPU and Memory**, and select **MetaStore Memory Usage Statistics** and click **OK**, check whether the used heap memory of the MetaStore service reaches the threshold(default value: 95%) of the maximum heap memory specified for MetaStore.

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Hive** > **Configurations > All Configurations**. Choose **HiveServer/MetaStore** > **JVM**. Adjust the value of **-Xmx** in **HIVE_GC_OPTS/ METASTORE_GC_OPTS** as the following rules. Click **Save**.

NOTE

Suggestions for GC parameter settings for the HiveServer:

- When the heap memory used by the HiveServer process reaches the threshold (default value: 95%) of the maximum heap memory set by the HiveServer process, change the value of **-Xmx** to twice the default value. For example, if **-Xmx** is set to 2GB by default, change the value of **-Xmx** to 4GB. You are advised to change the value of **-Xms** to set the ratio of **-Xms** and **-Xmx** to 1:2 to avoid performance problems when JVM dynamically. On the FusionInsight Manager home page, choose **O&M**> **Alarm**> **Thresholds** > *Name of the desired cluster* **>** **Hive** > **CPU and Memory** > **HiveServer Heap Memory Usage Statistics (HiveServer)** to view **Threshold**.

Suggestions for GC parameter settings for the MetaServer:

- When the heap memory used by the MetaStore process reaches the threshold (default value: 95%) of the maximum heap memory set by the MetaStore process, change the value of **-Xmx** to twice the default value. For example, if **-Xmx** is set to 2GB by default, change the value of **-Xmx** to 4GB. On the FusionInsight Manager home page, choose **O&M**> **Alarm**> **Thresholds** > *Name of the desired cluster* **>** **Hive** > **CPU and Memory** > **MetaStore Heap Memory Usage Statistics (MetaStore)** to view **Threshold**.
- You are advised to change the value of **-Xms** to set the ratio of **-Xms** and **-Xmx** to 1:2 to avoid performance problems when JVM dynamically.

Step 5 Click **More > Restart Service** to restart the service.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **Hive** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.115 ALM-16006 The Direct Memory Usage of the Hive Process Exceeds the Threshold

Description

The system checks the Hive service status every 30 seconds. The alarm is generated when the direct memory usage of an Hive service exceeds the threshold (95% of the maximum memory).

Users can choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Hive** to change the threshold.

The alarm is cleared when the direct memory usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
16006	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

When the direct memory usage of Hive is overhigh, the performance of Hive task operation is affected. In addition, a memory overflow may occur so that the Hive service is unavailable.

Possible Causes

The direct memory of the Hive instance on the node is overused or the direct memory is inappropriately allocated. As a result, the usage exceeds the threshold.

Procedure

Check direct memory usage.

Step 1 On the FusionInsight Manager portal, click **O&M > Alarm > Alarms** and select the alarm whose **Alarm ID** is **16006**. Then check the role name in **Location** and confirm the IP adress of the instance.

- If the role for which the alarm is generated is HiveServer, go to **Step 2**.
- If the role for which the alarm is generated is MetaStore, go to **Step 3**.

Step 2 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services > Hive > Instance** and click the HiveServer for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **CPU and Memory**, and select **HiveServer Memory Usage Statistics** and click **OK**, check whether the used direct memory of the HiveServer service reaches the threshold(default value: 95%) of the maximum direct memory specified for HiveServer.

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 3 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Hive** > **Instance** and click the MetaStore for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **CPU and Memory**, and select **MetaStore Memory Usage Statistics** and click **OK**, check whether the used direct memory of the MetaStore service reaches the threshold(default value: 95%) of the maximum direct memory specified for MetaStore.

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 4 On the FusionInsight Manager portal, choose **Cluster** >*Name of the desired cluster* > **Services** > **Hive** > **Configurations > All Configurations**. Choose **HiveServer/MetaStore** > **JVM**. Adjust the value of **-XX:MaxDirectMemorySize** in **HIVE_GC_OPTS/METASTORE_GC_OPTS** as the following rules. Click **Save**.

NOTE

Suggestions for GC parameter settings for the HiveServer:

- It is recommended that you set the value of **-XX:MaxDirectMemorySize** to 1/8 of the value of **-Xmx**. For example, if **-Xmx** is set to 8 GB, **-XX:MaxDirectMemorySize** is set to 1024 MB. If **-Xmx** is set to 4 GB, **-XX:MaxDirectMemorySize** is set to 512 MB. It is recommended that the value of **-XX:MaxDirectMemorySize** be greater than or equal to 512 MB.

Suggestions for GC parameter settings for the MetaServer:

- It is recommended that you set the value of **-XX:MaxDirectMemorySize** to 1/8 of the value of **-Xmx**. For example, if **-Xmx** is set to 8 GB, **-XX:MaxDirectMemorySize** is set to 1024 MB. If **-Xmx** is set to 4 GB, **-XX:MaxDirectMemorySize** is set to 512 MB. It is recommended that the value of **-XX:MaxDirectMemorySize** be greater than or equal to 512 MB.

Step 5 Click **More > Restart Service** to restart the service.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **Hive** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected fault logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.116 ALM-16007 Hive GC Time Exceeds the Threshold

Description

The system checks the garbage collection (GC) time of the Hive service every 60 seconds. This alarm is generated when the detected GC time exceeds the threshold (exceeds 12 seconds for three consecutive checks.) To change the threshold, choose **O&M > Alarm > Thresholds >** *Name of the desired cluster* > **Hive**. This alarm is cleared when the Hive GC time is shorter than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
16007	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the GC time exceeds the threshold, Hive data read and write are affected.

Possible Causes

The memory of Hive instances is overused, the heap memory is inappropriately allocated. As a result, GCs occur frequently.

Procedure

Check the GC time.

Step 1 On the FusionInsight Manager portal, click **O&M > Alarm > Alarms** and select the alarm whose **Alarm ID** is **16007**. Then check the role name in **Location** and confirm the IP adress of the instance.

- If the role for which the alarm is generated is HiveServer, go to **Step 2**.
- If the role for which the alarm is generated is MetaStore, go to **Step 3**.

Step 2 On the FusionInsight Manager portal, choose **Cluster** >*Name of the desired cluster* > **Services** > **Hive** > **Instance** and click the HiveServer for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **GC**, and select **Garbage Collection (GC) Time of HiveServer** and click **OK** to check whether the GC time is longer than 12 seconds.

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 3 On the FusionInsight Manager portal, choose **Cluster** >*Name of the desired cluster* > **Services** > **Hive** > **Instance** and click the MetaStore for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **GC**, and select **Garbage Collection (GC) Time of MetaStore** and click **OK** to check whether the GC time is longer than 12 seconds.

- If yes, go to **Step 4**.

- If no, go to **Step 7**.

Check the current JVM configuration.

Step 4 On the FusionInsight Manager portal, choose **Cluster** >*Name of the desired cluster* > **Services > Hive > Configurations > All Configurations**. Choose **HiveServer/MetaStore** > **JVM**. Adjust the value of **-Xmx** in **HIVE_GC_OPTS/ METASTORE_GC_OPTS** as the following rules. Click **Save**.

NOTE

Suggestions for GC parameter settings for the HiveServer:

- When the Hive GC time exceeds the threshold, change the value of **-Xmx** to twice the default value. For example, if **-Xmx** is set to 2 GB by default, change the value of **-Xmx** to 4 GB.
- You are advised to change the value of **-Xms** to set the ratio of **-Xms** and **-Xmx** to 1:2 to avoid performance problems when JVM dynamically.

Suggestions for GC parameter settings for the MetaServer:

- When the Meta GC time exceeds the threshold, change the value of **-Xmx** to twice the default value. For example, if **-Xmx** is set to 2 GB by default, change the value of **-Xmx** to 4 GB.
- You are advised to change the value of **-Xms** to set the ratio of **-Xms** and **-Xmx** to 1:2 to avoid performance problems when JVM dynamically.

Step 5 Click **More > Restart Service** to restart the service.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal of active and standby clusters, choose **O&M** > **Log > Download**.

Step 8 In the **Service**, select **Hive** in the required cluster.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.117 ALM-16008 Non-Heap Memory Usage of the Hive Process Exceeds the Threshold

Description

The system checks the Hive service status every 30 seconds. The alarm is generated when the non-heap memory usage of an Hive service exceeds the threshold (95% of the maximum memory).

Users can choose **O&M > Alarm > Thresholds >** *Name of the desired cluster* > **Hive** to change the threshold.

The alarm is cleared when the non-heap memory usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
16008	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.

Impact on the System

When the non-heap memory usage of Hive is overhigh, the performance of Hive task operation is affected. In addition, a memory overflow may occur so that the Hive service is unavailable.

Possible Causes

The non-heap memory of the Hive instance on the node is overused or the non-heap memory is inappropriately allocated. As a result, the usage exceeds the threshold.

Procedure

Check non-heap memory usage.

Step 1 On the FusionInsight Manager portal, click **O&M > Alarm > Alarms** and select the alarm whose **Alarm ID** is **16008**. Then check the role name in **Location** and confirm the IP adress of the instance.

- If the role for which the alarm is generated is HiveServer, go to **Step 2**.
- If the role for which the alarm is generated is MetaStore, go to **Step 3**.

Step 2 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Hive** > **Instance** and click the HiveServer for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **CPU and Memory**, and select **HiveServer Memory Usage Statistics** and click **OK**, check whether the used non-heap memory of the HiveServer service reaches the threshold(default value: 95%) of the maximum non-heap memory specified for HiveServer.

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 3 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Hive** > **Instance** and click the MetaStore for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **CPU and Memory**, and select **MetaStore Memory Usage Statistics** and click **OK**, check whether the used non-heap memory of the MetaStore service reaches the threshold(default value: 95%) of the maximum non-heap memory specified for MetaStore.

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 4 On the FusionInsight Manager portal, choose **Cluster > ***Name of the desired cluster* > **Services** > **Hive** > **Configurations > All Configurations**. Choose **HiveServer/MetaStore** > **JVM**. Adjust the value of **-XX:MaxMetaspaceSize** in **HIVE_GC_OPTS/METASTORE_GC_OPTS** as the following rules. Click **Save**.

NOTE

Suggestions for GC parameter settings for the HiveServer:

- It is recommended that you set the value of **-XX:MaxMetaspaceSize** to 1/8 of the value of **-Xmx**. For example, if **-Xmx** is set to 2 GB, **-XX:**

  MaxMetaspaceSize is set to 256 MB. If **-Xmx** is set to 4 GB, **-XX:MaxMetaspaceSize** is set to 512 MB.

Suggestions for GC parameter settings for the MetaServer:

- It is recommended that you set the value of **-XX:MaxMetaspaceSize** to 1/8 of the value of **-Xmx**. For example, if **-Xmx** is set to 2 GB, **-XX:**

  MaxMetaspaceSize is set to 256 MB. If **-Xmx** is set to 4 GB, **-XX:MaxMetaspaceSize** is set to 512 MB

Step 5 Click **More > Restart Service** to restart the service.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **Hive** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.118 ALM-16009 Map Number Exceeds the Threshold

Description

The system checks the number of HQL maps in every 30 seconds. This alarm is generated if the number exceeds the threshold. By default, **Trigger Count** is set to **3**, and the threshold is 5000.

Attribute

Alarm ID	Alarm Severity	Auto Clear
16009	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Name	Meaning
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the number of HQL maps executed on Hive is excessively large, the HQL execution speed is slow, and a large number of resources are occupied.

Possible Causes

The HQL statements are not the optimal.

Procedure

Check the number of HQL maps.

Step 1 On FusionInsight Manager portal, choose **Cluster** >*Name of the desired cluster* > **Services** > **Hive** > **Resource**. Check the HQL statements with the excessively large number (5000 or more) of maps in **HQL Map Count**.

Step 2 Locate the corresponding HQL statements, optimize them and execute them again.

Step 3 Check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 4**.

Collect fault information.

Step 4 On the FusionInsight Manager, choose **O&M** > **Log > Download**.

Step 5 Select **Hive** in the required cluster from the **Service**.

Step 6 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 7 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.119 ALM-16045 Hive Data Warehouse Is Deleted

Description

The system checks the Hive data warehouse in every 60 seconds.This alarm is generated when the Hive data warehouse is deleted.

Attribute

Alarm ID	Alarm Severity	Auto Clear
16045	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The default Hive data warehouse is deleted. As a result, creating databases or tables in the default data warehouse fails, and services are affected.

Possible Causes

Hive periodically checks the status of the default data warehouse and finds that the default data warehouse is deleted.

Procedure

Check the default Hive data warehouse.

Step 1 Log in to the node where the client is located as user **root**.

Step 2 Run the following command to check whether the **warehouse** directory exists in **hdfs://hacluster/user/***<username>***/.Trash/Current/**.

hdfs dfs -ls hdfs://hacluster/user/*<username>***/.Trash/Current/**

For example, if **user/hive/warehouse** exists:

```
host01:/opt/client # hdfs dfs -ls hdfs://hacluster/user/test/.Trash/Current/
Found 1 items
drwx------   - test hadoop          0 2019-06-17 19:53 hdfs://hacluster/user/test/.Trash/Current/user
```

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 By default, there is an automatic recovery mechanism for the data warehouse. You can wait for 5 ~10s to check whether the default data warehouse is restored. If the data warehouse is not recovered, manually run the following command to restore the data warehouse.

hdfs dfs -mv hdfs://hacluster/user/*<username>***/.Trash/Current/user/hive/warehouse /user/hive/warehouse**

Step 4 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Collect fault information.

Step 5 Collect related information in the **.Trash/Current/** directory on the client background.

Step 6 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.120 ALM-16046 Hive Data Warehouse Permission Is Modified

Description

The system checks the Hive data warehouse permission in every 60 seconds. This alarm is generated if the permission is modified.

Attribute

Alarm ID	Alarm Severity	Auto Clear
16046	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

If the permission on the Hive default data warehouse is modified, the permission for users or user groups to create databases or tables in the default data warehouse is changed.

Possible Causes

Hive periodically checks the status of the default data warehouse and finds that default data warehouse permission is changed.

Procedure

Check the Hive default data warehouse permission.

Step 1 Log in to the node where the client is located as user **root**.

Step 2 Run the following command to go to the HDFS client installation directory:

cd *Client installation directory*

source bigdata_env

kinit *User who has the supergroup permission* (Skip this step for a common cluster.)

Step 3 Run the following command to restore the default data warehouse permission:

- Security mode: **hdfs dfs -chmod 770 hdfs://hacluster/user/hive/warehouse**
- Non-security mode: **hdfs dfs -chmod 777 hdfs://hacluster/user/hive/warehouse**

Step 4 Check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 5**.

Collect fault information.

Step 5 Collect related information in the **hdfs://hacluster/user/hive/warehouse** directory on the client background.

Step 6 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.121 ALM-16047 HiveServer Has Been Deregistered from ZooKeeper

Description

The system checks the Hive service every 60 seconds. This alarm is generated when Hive registration information on ZooKeeper is lost or Hive cannot connect to ZooKeeper.

Attribute

Alarm ID	Alarm Severity	Auto Clear
16047	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

If the Hive configuration cannot be read from ZooKeeper, HiveServer will be unavailable.

Possible Causes

- The network is disconnected.
- The ZooKeeper instance is abnormal.

Procedure

Restart related instances.

Step 1 Log in to FusionInsight Manager. Choose **O&M** > **Alarm** > **Alarms**, click the drop-down list in the row that contains the alarm, and view role and the IP address of the node for which the alarm is generated in **Location**.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Hive** > **Instance**, select the instance at the IP address for which the alarm is generated, and choose **More** > **Restart Instance**.

Step 3 Wait for 5 minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Collect the fault information.

Step 4 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 5 Expand the **Service** drop-down list, and select **Hive** for the target cluster.

Step 6 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 7 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.122 ALM-16048 Tez or Spark Library Path Does Not Exist

Description

The system checks the Tez and Spark library paths every 180 seconds. This alarm is generated when the Tez or Spark library path does not exist.

Attribute

Alarm ID	Alarm Severity	Auto Clear
16048	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The Hive on Tez and Hive on Spark functions are affected.

Possible Causes

The Tez or Spark library path is deleted from the HDFS.

Procedure

Check the default Hive data warehouse.

Step 1 Log in to the node where the client is located as user **root**.

Step 2 Run the following command to check whether the **tezlib** or **sparklib** directory exists in the **hdfs://hacluster/user/{User name}/.Trash/Current/** director:

hdfs dfs -ls hdfs://hacluster/user/*<username>***/.Trash/Current/**

For example, the following information shows that **/user/hive/tezlib/8.1.0.1/** and **/user/hive/sparklib/8.1.0.1/** exist.

```
host01:/opt/client # hdfs dfs -ls hdfs://hacluster/user/test/.Trash/Current/
Found 1 items
drwx------   - test hadoop          0 2019-06-17 19:53 hdfs://hacluster/user/test/.Trash/Current/user
```

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 Run the following command to restore **tezlib** and **sparklib**.

hdfs dfs -mv hdfs://hacluster/user/*<username>***/.Trash/Current/user/hive/tezlib/8.1.0.1/tez.tar.gz /user/hive/tezlib/8.1.0.1/tez.tar.gz**

Step 4 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Collect fault information.

Step 5 Collect related information in the **.Trash/Current/** directory on the client background.

Step 6 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.123 ALM-17003 Oozie Service Unavailable

Description

The system checks the Oozie service status in every 5 seconds. This alarm is generated when Oozie or a component on which Oozie depends cannot provide services properly.

This alarm is automatically cleared when the Oozie service recovers.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
17003	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

Oozie cannot be used to submit jobs.

Possible Causes

- The DBService service is abnormal or the data of Oozie stored in DBService is damaged.
- The HDFS service is abnormal or the data of Oozie stored in HDFS is damaged.
- The Yarn service is abnormal.
- The Nodeagent process is abnormal.

Procedure

Query the Oozie service health status code.

Step 1 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* >**Services** > **Oozie**. Click **oozie** (any one is OK) on the **oozie WebUI**. to go to the Oozie WebUI.

NOTE

By default, the **admin** user does not have the permissions to manage other components. If the page cannot be opened or the displayed content is incomplete when you access the native UI of a component due to insufficient permissions, you can manually create a user with the permissions to manage that component.

Step 2 Add **/servicehealth** to the URL in the address box of the browser and access again. The value of **statusCode** is the current Oozie service health status code.

For example, visit **https://10.10.0.117:20026/Oozie/oozie/130/oozie/servicehealth**. The result is as follows:

```
{"beans":[{"name":"serviceStatus","statusCode":0}]}
```

If the health status code cannot be displayed or the browser does not respond, the service may be unavailable due to Oozie process fault. See **Step 13** to rectify the fault.

Step 3 Perform the operations based on the error code. For details, see **Table 9-92**.

Table 9-92 Oozie service health status code

Status Code	Description	Error Cause	Solution
0	The service is running properly.	None	None
18002	The DBService service is abnormal.	Oozie fails to connect to DBService or the data stored in DBService is damaged.	See **Step 4**.

Status Code	Description	Error Cause	Solution
18003	The HDFS service is abnormal.	Oozie fails to connect to HDFS or the data stored in HDFS is damaged.	See **Step 7**.
18005	The MapReduce service is abnormal.	The Yarn service is abnormal.	See **Step 11**.

Check the DBService service.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services**, and check whether the DBService service is running properly.

- If yes, go to **Step 6**.
- If no, go to **Step 5**.

Step 5 Resolve the problem of DBService based on the alarm help and check whether the Oozie alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 18**.

Step 6 Log in to the Oozie database to check whether the data is complete.

1. Log in to the active DBService node as user **root**.

   On the FusionInsight Manager page, choose **Cluster** > *Name of the desired cluster* > **Services** > **DBService > Instance** to view the IP address of the active DBservice node.

2. Run the following command to log in to the Oozie database:

   su - omm

   source ${BIGDATA_HOME}/FusionInsight_BASE_8.1.0.1/install/FusionInsight-dbservice-2.7.0/.dbservice_profile

   gsql -U *Username* **-W** *Oozie database password* **-p 20051 -d** *Database name*

3. After the login is successful, enter **\d** to check whether there are 15 data tables.

   The Oozie service has 15 data tables by default. If these data tables are deleted or the table structure is modified, the Oozie service may be unavailable. Contact the O&M personnel to back up the data and perform restoration.

Check the HDFS service.

Step 7 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services**, and check whether the HDFS service is running properly.

- If yes, go to **Step 9**.
- If no, go to **Step 8**.

Step 8 Resolve the problem of HDFS based on the alarm help and check whether the Oozie alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 18**.

Step 9 Log in to HDFS to check whether the Oozie file directory structure is complete.

1. Download and install an HDFS client..
2. Log in to the client node as user **root** and run the following commands to check whether **/user/oozie/share** exists.

   If the cluster uses the security mode, perform security authentication.

   kinit admin

   hdfs dfs -ls /user/oozie/share

- If yes, go to **Step 18**.
- If no, go to **Step 10**.

Step 10 In the Oozie client installation directory, manually upload the share directory to **/user/oozie** in HDFS, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 18**.

Check the Yarn and MapReduce service.

Step 11 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services**, and check whether the Yarn and MapReduce services are running properly.

- If yes, go to **Step 18**.
- If no, go to **Step 12**.

Step 12 Resolve the problem of Yarn and MapReduce based on the alarm help and check whether the Oozie alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 18**.

Check the Oozie process.

Step 13 Log in to each node of Oozie as user **root**.

Step 14 Run the **ps -ef | grep oozie** command to check whether the Oozie process exists.

- If yes, go to **Step 15**.
- If no, go to **Step 18**.

Step 15 Collect fault information in **prestartDetail.log**, **oozie.log**, and **catalina.out** in the Oozie log directory **/var/log/Bigdata/oozie**. If the alarm is not caused by manual misoperation, go to **Step 16**.

Check the Nodeagent process.

Step 16 Log in to each node of Oozie as user **root**. Run the **ps -ef | grep nodeagent** command to check whether the Nodeagent process exists.

- If yes, go to **Step 17**.
- If no, go to **Step 18**.

Step 17 Run the **kill -9** *The process ID of nodeagent* command, wait 10 minutes, and check whether alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 18**.

Step 18 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.124 ALM-17004 Oozie Heap Memory Usage Exceeds the Threshold

Description

The system checks the heap memory usage of the Oozie service every 60 seconds. The alarm is generated when the heap memory usage of a Metadata instance exceeds the threshold (95% of the maximum memory). The alarm is cleared when the heap memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
17004	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

The heap memory overflow may cause a service breakdown.

Possible Causes

The heap memory of the Oozie instance is overused or the heap memory is inappropriately allocated.

Procedure

Check heap memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** > **Oozie Heap Memory Usage Exceeds the Threshold** > **Location**. Check the IP address of the instance involved in this alarm.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* > **Services** > **Oozie** > **Instance**. Click the instance for which the alarm is generated to go to the page for the instance. Click the drop-down menu in the chart area and choose **Customize** > **Memory** > **Oozie Heap Memory Resource Percentage**. Click **OK**.

Step 3 Check whether the used heap memory of Oozie reaches the threshold (the default value is 95% of the maximum heap memory) specified for Oozie.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* > **Services** > **Oozie** > **Configurations > All Configurations**. Set Search **GC_OPTS** in the search box. Increase the value of **-Xmx** as required, and click **Save** > **OK**.

NOTE

Suggestions on GC parameter settings for Oozie:

You are advised to set **-Xms** and **-Xmx** to the same value to prevent adverse impact on performance when JVM dynamically adjusts the heap memory size.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 7 Select **Oozie** in the required cluster from the **Service**.

Step 8 Click in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.125 ALM-17005 Oozie Non Heap Memory Usage Exceeds the Threshold

Description

The system checks the non heap memory usage of Oozie every 30 seconds. This alarm is reported if the non heap memory usage of Oozie exceeds the threshold (80%). This alarm is cleared if the non heap memory usage is lower than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
17005	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Non-heap memory overflow may cause service breakdown.

Possible Causes

The non-heap memory of the Oozie instance is overused or the non-heap memory is inappropriately allocated.

Procedure

Check non-heap memory usage.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **Oozie Non Heap Memory Usage Exceeds the Threshold**. On the displayed page, check the location information of the alarm. Check the name of the instance host for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > *Name of the target cluster* > **Services** > **Oozie** and click the **Instance** tab. On the displayed page, select the role corresponding to the host name for which the alarm is generated and select **Customize** from the drop-down list in the upper right corner of the chart area. Choose **Memory** and select **Oozie Non Heap Memory Resource Percentage**. Click **OK**.

Step 3 Check whether the non-heap memory used by Oozie reaches the threshold (80% of the maximum non-heap memory by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > *Name of the target cluster* > **Services** > **Oozie** and click the **Configurations** and then **All Configurations**. On the displayed page, search for the **GC_OPTS** parameter in the search box and check whether it contains **-XX: MaxMetaspaceSize**. If yes, increase the value of **-XX: MaxMetaspaceSize** based on the site requirements. If no, manually add **-XX: MaxMetaspaceSize** and set its value to 1/8 of the value of **-Xmx**. Click **Save**, and then click **OK**

NOTE

JDK1.8 does not support the **MaxPermSize** parameter.

Suggestions on GC parameter settings for Oozie:

Set the value of **-XX:MaxMetaspaceSize** to 1/8 of the value of **-Xmx**. For example, if **-Xmx** is set to 2 GB, **-XX:MaxMetaspaceSize** is set to 256 MB. If **-Xmx** is set to 4 GB, **-XX:MaxMetaspaceSize** is set to 512 MB.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Oozie** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.126 ALM-17006 Oozie Direct Memory Usage Exceeds the Threshold

Description

The system checks the direct memory usage of the Oozie service every 30 seconds. The alarm is generated when the direct memory usage of an Oozie instance exceeds the threshold (80% of the maximum memory). The alarm is cleared when the direct memory usage of Oozie is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
17006	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

The direct memory overflow may cause a service breakdown.

Possible Causes

The direct memory of the Oozie instance is overused or the direct memory is inappropriately allocated.

Procedure

Check direct memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms** > **Oozie Direct Memory Usage Exceeds the Threshold** > **Location**. Check the IP address of the instance involved in this alarm.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* > **Services** > **Oozie** > **Instance**. Click the instance for which the alarm is generated to go to the page for the instance. Click the drop-down menu in the chart area and choose **Customize** > **Memory** > **Oozie Direct Buffer Resource Percentage**. Click **OK**.

Step 3 Check whether the used direct memory of Oozie reaches the threshold (the default value is 80% of the maximum direct memory) specified for Oozie.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* > **Services** > **Oozie** > **Configurations**. Click **All Configurations**. Search **GC_OPTS** in the search box. Increase the value of **-XX:MaxDirectMemorySize** as required, and click **Save**. Click **OK**.

NOTE

Suggestions on GC parameter settings for Oozie:

You are advised to set the value of **-XX:MaxDirectMemorySize** to 1/4 of the value of **-Xmx**. For example, if **-Xmx** is set to 4 GB, **-XX:MaxDirectMemorySize** is set to 1024 MB. If **-Xmx** is set to 2 GB, **-XX:MaxDirectMemorySize** is set to 512 MB. It is recommended that the value of **-XX:MaxDirectMemorySize** be greater than or equal to 512 MB.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 7 Select **Oozie** in the required cluster from the **Service** drop-down list.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.127 ALM-17007 Garbage Collection (GC) Time of the Oozie Process Exceeds the Threshold

Description

The system checks GC time of the Oozie process every 60 seconds. The alarm is generated when GC time of the Oozie process exceeds the threshold (default value: **12 seconds**). The alarm is cleared when GC time is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
17007	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

Oozie responds slowly when it is used to submit tasks.

Possible Causes

The heap memory of the Oozie instance is overused or the heap memory is inappropriately allocated.

Procedure

Check GC time.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms** > **Garbage Collection (GC) Time of the Oozie Process Exceeds the Threshold** > **Location**. Check the IP address of the instance involved in this alarm.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* > **Services** > **Oozie** > **Instance**. Click the instance for which the alarm is generated to go to the page for the instance. Click the drop-down menu in the chart area and choose **Customize** > **GC** > **Garbage Collection (GC) Time of Oozie**. Click **OK**.

Step 3 Check whether GC time of the Oozie process every second exceeds the threshold (default value: **12 seconds**).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* > **Services** > **Oozie** > **Configurations**. Click **All Configurations**. Search **GC_OPTS** in the search box. Increase the value of **-Xmx** as required, and click **Save**. Click **OK**.

NOTE

Suggestions on GC parameter settings for Oozie:

You are advised to set **-Xms** and **-Xmx** to the same value to prevent adverse impact on performance when JVM dynamically adjusts the heap memory size.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 7 Select **Oozie** in the required cluster from the **Service** drop-down list.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.128 ALM-18000 Yarn Service Unavailable

Description

This alarm is generated when the Yarn service is unavailable. The alarm module checks the Yarn service status every 60 seconds.

The alarm is cleared when the Yarn service recovers.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18000	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceNam	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The cluster cannot provide Yarn services. Users cannot run new applications. Submitted applications cannot be run.

Possible Causes

- The ZooKeeper service is abnormal.
- The HDFS service is abnormal.
- There is no active ResourceManager instance in the Yarn cluster.
- All the NodeManagers in the Yarn cluster are abnormal.

Procedure

Check ZooKeeper service status.

Step 1 On the FusionInsight Manager, check whether the alarm list contains **ALM-13000 ZooKeeper Service Unavailable**.

- If yes, go to **Step 2**.
- If no, go to **Step 3**.

Step 2 Rectify the fault by following the steps provided in **ALM-13000 ZooKeeper Service Unavailable**, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 3**.

Check the HDFS service status.

Step 3 On the FusionInsight Manager, check whether the alarm list contains the HDFS alarms.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 Choose **O&M > Alarm > Alarms**, handle HDFS alarms based on the alarm help, and check whether the Yarn alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check the ResourceManager status in the Yarn cluster.

Step 5 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn**.

Step 6 In **Dashboard**, check whether there is an active ResourceManager instance in the Yarn cluster.

- If yes, go to **Step 7**.
- If no, go to **Step 10**.

Check the NodeManager node status in the Yarn cluster.

Step 7 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn** > **Instance**.

Step 8 Query NodeManager **Running Status**, and check whether there are unhealthy nodes.

- If yes, go to **Step 9**.
- If no, go to **Step 10**.

Step 9 Rectify the fault by following the steps provided in **ALM-18002 NodeManager Heartbeat Lost** or **ALM-18003 NodeManager Unhealthy**. After the fault is rectified, check whether the Yarn alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 10**.

Collect fault information.

Step 10 On the FusionInsight Manager portal of the active cluster, choose **O&M** > **Log > Download**.

Step 11 Select **Yarn** in the required cluster from the **Service**.

Step 12 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 13 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.129 ALM-18002 NodeManager Heartbeat Lost

Description

The system checks the number of lost NodeManager nodes every 30 seconds, and compares the number with the threshold. The Number of Lost Nodes indicator has a default threshold. The alarm is generated when the value of Number of Lost Nodes exceeds the threshold.

To change the threshold, on FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn**. On the displayed page, choose **Configurations** > **All Configurations**, and change the value of **yarn.nodemanager.lost.alarm.threshold**. You do not need to restart Yarn to make the change take effect.

The default threshold is 0. The alarm is generated when the number of lost nodes exceeds the threshold, and is cleared when the number of lost nodes is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18002	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Lost Host	Specifies the list of hosts with lost nodes.

Impact on the System

- The lost NodeManager node cannot provide the Yarn service.
- The number of containers decreases, so the cluster performance deteriorates.

Possible Causes

- NodeManager is forcibly deleted without decommission.
- All the NodeManager instances are stopped or the NodeManager process is faulty.
- The host where the NodeManager node resides is faulty.
- The network between the NodeManager and ResourceManager is disconnected or busy.

Procedure

Check the NodeManager status.

Step 1 On the FusionInsight Manager, and choose **O&M** > **Alarm > Alarms**. Click ⌄ before the alarm and obtain lost nodes in **Additional Information**.

Step 2 Check whether the lost nodes are hosts that have been manually deleted without decommission.

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 After the setting, Choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn**. On the displayed page, choose **Configurations** > **All Configurations**. Search for **yarn.nodemanager.lost.alarm.threshold** and change its value to the number of hosts that are not out of service and proactively deleted. After the setting, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Step 4 Manually clear the alarm. Note that decommission must be performed before deleting hosts.

Step 5 On the FusionInsight Manager portal, choose **Cluster > Hosts**, and check whether the nodes obtained in **Step 1** are healthy.

- If yes, go to **Step 7**.
- If no, go to **Step 6**.

Step 6 Rectify the node fault based on **ALM-12006 Node Fault** and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Check the process status.

Step 7 On the FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* **> Services** > **Yarn** > **Instance**, and check whether there are NodeManager instances whose status is not **Good**.

- If yes, go to **Step 10**.
- If no, go to **Step 8**.

Step 8 Check whether the NodeManager instance is deleted.

- If yes, go to **Step 9**.
- If no, go to **Step 11**.

Step 9 Restart the active and standby ResourceManager instances, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 13**.

Check the instance status.

Step 10 Select NodeManager instances which running state is not **Normal** and restart them. Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Check the network status.

Step 11 Log in to the management node, **ping** the IP address of the lost NodeManager node to check whether the network is disconnected or busy.

- If yes, go to **Step 12**.
- If no, go to **Step 13**.

Step 12 Rectify the network, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 13**.

Collect fault information.

Step 13 On the FusionInsight Manager in the active cluster, choose **O&M** > **Log > Download**.

Step 14 Select **Yarn** in the required cluster from the **Service**.

Step 15 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 16 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.130 ALM-18003 NodeManager Unhealthy

Description

The system checks the number of unhealthy NodeManager nodes every 30 seconds, and compares the number with the threshold. The Unhealthy Nodes indicator has a default threshold. This alarm is generated when the value of the Unhealthy Nodes indicator exceeds the threshold.

To change the threshold, on FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn**. On the displayed page, choose **Configurations** > **All Configurations**, and change the value of **yarn.nodemanager.unhealthy.alarm.threshold**. You do not need to restart Yarn to make the change take effect.

The default threshold is 0. The alarm is generated when the number of unhealthy nodes exceeds the threshold, and is cleared when the number of unhealthy nodes is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18003	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.

Name	Meaning
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Unhealthy Host	Specifies the list of hosts with unhealthy nodes.

Impact on the System

- The faulty NodeManager node cannot provide the Yarn service.
- The number of containers decreases, so the cluster performance deteriorates.

Possible Causes

- The hard disk space of the host where the NodeManager node resides is insufficient.
- User **omm** does not have the permission to access a local directory on the NodeManager node.

Procedure

Check the hard disk space of the host.

Step 1 On the FusionInsight Manager, and choose **O&M** > **Alarm > Alarms**. Click ⌄ before the alarm and obtain unhealthy nodes in **Additional Information**.

Step 2 Choose **Cluster** > *Name of the desired cluster* **> Services > Yarn** > **Instance**, select the NodeManager instance corresponding to the host, choose **Instance Configurations > All Configurations** and view disks corresponding to **yarn.nodemanager.local-dirs** and **yarn.nodemanager.log-dirs**.

Step 3 Choose **O&M > Alarm > Alarms**. In the alarm list, check whether the related disk has the alarm **ALM-12017 Insufficient Disk Capacity**.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 Rectify the disk fault based on **ALM-12017 Insufficient Disk Capacit** and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Step 5 Choose **Hosts** > *Name of the desired host* . On the **Dashboard** page, check the disk usage of the corresponding partition. Check whether the percentage of the used space of the mounted disk exceeds the value of **yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage**

- If yes, go to **Step 6**.

- If no, go to **Step 7**.

Step 6 Reduce the disk usage to less than the value of **yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage**, wait for 10 to 20 minutes, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Check the access permission of the local directory on each NodeManager node.

Step 7 Obtain the NodeManager directory viewed in **Step 2**, log in to each NodeManager node as user **root**, and go to the obtained directory.

Step 8 Run the **ll** command to check whether the permission of the **localdir** and **containerlogs** folders is **755** and whether **User:Group** is **omm:ficommon**.

- If yes, no further action is required.
- If no, go to **Step 9**.

Step 9 Run the following command to set the permission to **755** and **User:Group** to **omm:ficommon**:

chmod 755 *<folder_name>*

chown omm:ficommon *<folder_name>*

Step 10 Wait for 10 to 20 minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Collect fault information.

Step 11 On the FusionInsight Manager in the active cluster, choose **O&M** > **Log > Download**.

Step 12 Select **Yarn** in the required cluster from the **Service**.

Step 13 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 14 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.131 ALM-18008 Heap Memory Usage of ResourceManager Exceeds the Threshold

Description

The system checks the heap memory usage of Yarn ResourceManager every 30 seconds and compares the actual usage with the threshold. The alarm is generated when the heap memory usage of Yarn ResourceManager exceeds the threshold (95% of the maximum memory by default).

Users can choose **O&M > Alarm > Thresholds** > *Name of the desired cluster* > **Yarn** to change the threshold.

When the **Trigger Count** is 1, this alarm is cleared when the heap memory usage of Yarn ResourceManager is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the heap memory usage of Yarn ResourceManager is less than or equal to 95% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18008	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

When the heap memory usage of Yarn ResourceManager is overhigh, the performance of Yarn task submission and operation is affected. In addition, a memory overflow may occur so that the Yarn service is unavailable.

Possible Causes

The heap memory of the Yarn ResourceManager instance on the node is overused or the heap memory is inappropriately allocated. As a result, the usage exceeds the threshold.

Procedure

Check the heap memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** > **Heap Memory Usage of Yarn ResourceManager Exceeds the Threshold** > **Location**. Check the HostName of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* **> Services** > **Yarn** > **Instance** > **ResourceManager** (Indicates the host name of the instance for which the alarm is generated)**.** Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **ResourceManager > Percentage of Used Memory of the ResourceManager**. Check the heap memory usage.

Step 3 Check whether the used heap memory of ResourceManager reaches 95% of the maximum heap memory specified for ResourceManager.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* **> Services** > **Yarn** > **Configurations** > **All Configurations** > **ResourceManager** > **System**. Increase the value of **GC_OPTS** parameter as required, click **Save**. Restart the role instance.

NOTE

The mapping between the number of NodeManager instances in a cluster and the memory size of ResourceManager is as follows:

- If the number of NodeManager instances in the cluster reaches 100, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms4G -Xmx4G -XX:NewSize=512M -XX:MaxNewSize=1G
- If the number of NodeManager instances in the cluster reaches 200, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms6G -Xmx6G -XX:NewSize=512M -XX:MaxNewSize=1G
- If the number of NodeManager instances in the cluster reaches 500, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms10G -Xmx10G -XX:NewSize=1G -XX:MaxNewSize=2G
- If the number of NodeManager instances in the cluster reaches 1000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms20G -Xmx20G -XX:NewSize=1G -XX:MaxNewSize=2G
- If the number of NodeManager instances in the cluster reaches 2000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms40G -Xmx40G -XX:NewSize=2G -XX:MaxNewSize=4G
- If the number of NodeManager instances in the cluster reaches 3000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms60G -Xmx60G -XX:NewSize=2G -XX:MaxNewSize=4G
- If the number of NodeManager instances in the cluster reaches 4000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms80G -Xmx80G -XX:NewSize=2G -XX:MaxNewSize=4G
- If the number of NodeManager instances in the cluster reaches 5000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms100G -Xmx100G -XX:NewSize=3G -XX:MaxNewSize=6G

Step 5 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 7 Select the following node in the required cluster from the **Service**.

- NodeAgent
- Yarn

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.132 ALM-18009 Heap Memory Usage of JobHistoryServer Exceeds the Threshold

Description

The system checks the heap memory usage of Mapreduce JobHistoryServer every 30 seconds and compares the actual usage with the threshold. The alarm is generated when the heap memory usage of Mapreduce JobHistoryServer exceeds the threshold (95% of the maximum memory by default).

Users can choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Mapreduce** to change the threshold.

When the **Trigger Count** is 1, this alarm is cleared when the heap memory usage of MapReduce JobHistoryServer is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the heap memory usage of MapReduce JobHistoryServer is less than or equal to 95% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18009	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

When the heap memory usage of Mapreduce JobHistoryServer is overhigh, the performance of Mapreduce log archiving is affected. In addition, a memory overflow may occur so that the Yarn service is unavailable.

Possible Causes

The heap memory of the Mapreduce JobHistoryServer instance on the node is overused or the heap memory is inappropriately allocated. As a result, the usage exceeds the threshold.

Procedure

Check the memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** > **ALM-18009 Heap Memory Usage of MapReduce JobHistoryServer Exceeds the Threshold** > **Location**. Check the HostName of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* **> Services** > **Mapreduce** > **Instance** > **JobHistoryServer.** Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **JobHistoryServer heap memory usage statistics**. JobHistoryServer indicates the corresponding HostName of the instance for which the alarm is generated. Check the heap memory usage.

Step 3 Check whether the used heap memory of JobHistoryServer reaches 95% of the maximum heap memory specified for JobHistoryServer.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* **> Services** > **Mapreduce** > **Configurations** > **All Configurations** > **JobHistoryServer** > **System**. Increase the value of **GC_OPTS** parameter as required, click **Save**. Click **OK** and restart the role instance.

NOTE

The mapping between the number of historical tasks (10000) and the memory of JobHistoryServer is as follows:

-Xms30G -Xmx30G -XX:NewSize=1G -XX:MaxNewSize=2G

Step 5 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 7 Select the following node in the required cluster from the **Service**.

- NodeAgent
- Mapreduce

Step 8 Click in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.133 ALM-18010 ResourceManager GC Time Exceeds the Threshold

Description

The system checks the garbage collection (GC) duration of the ResourceManager process every 60 seconds. This alarm is generated when the GC duration exceeds the threshold (12 seconds by default).

This alarm is cleared when the GC duration is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18010	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

A long GC duration of the ResourceManager process may interrupt the services.

Possible Causes

The heap memory of the ResourceManager instance is overused or the heap memory is inappropriately allocated. As a result, GCs occur frequently.

Procedure

Check the GC duration.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** > **ALM-18010 ResourceManager GC Time Exceeds the Threshold** > **Location** to check the IP address of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* **>** **Services** > **Yarn** > **Instance** > **ResourceManager (IP address for which the alarm is generated)**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **Garbage Collection (GC) Time of ResourceManager** to check the GC duration statistics of the Broker process collected every minute.

Step 3 Check whether the GC duration of the ResourceManager process collected every minute exceeds the threshold (12 seconds by default).

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn** > **Configurations** > **All Configurations** > **ResourceManager** > **System** to increase the value of **GC_OPTS** parameter as required.

NOTE

The mapping between the number of NodeManager instances in a cluster and the memory size of ResourceManager is as follows:

- If the number of NodeManager instances in the cluster reaches 100, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms4G -Xmx4G -XX:NewSize=512M -XX:MaxNewSize=1G
- If the number of NodeManager instances in the cluster reaches 200, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms6G -Xmx6G -XX:NewSize=512M -XX:MaxNewSize=1G
- If the number of NodeManager instances in the cluster reaches 500, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms10G -Xmx10G -XX:NewSize=1G -XX:MaxNewSize=2G
- If the number of NodeManager instances in the cluster reaches 1000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms20G -Xmx20G -XX:NewSize=1G -XX:MaxNewSize=2G
- If the number of NodeManager instances in the cluster reaches 2000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms40G -Xmx40G -XX:NewSize=2G -XX:MaxNewSize=4G
- If the number of NodeManager instances in the cluster reaches 3000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms60G -Xmx60G -XX:NewSize=2G -XX:MaxNewSize=4G
- If the number of NodeManager instances in the cluster reaches 4000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms80G -Xmx80G -XX:NewSize=2G -XX:MaxNewSize=4G
- If the number of NodeManager instances in the cluster reaches 5000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms100G -Xmx100G -XX:NewSize=3G -XX:MaxNewSize=6G

Step 5 Save the configuration and restart the ResourceManager instance.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **ResourceManager** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.134 ALM-18011 NodeManager GC Time Exceeds the Threshold

Description

The system checks the garbage collection (GC) duration of the NodeManager process every 60 seconds. This alarm is generated when the GC duration exceeds the threshold (12 seconds by default).

This alarm is cleared when the GC duration is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18011	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

A long GC duration of the NodeManager process may interrupt the services.

Possible Causes

The heap memory of the NodeManager instance is overused or the heap memory is inappropriately allocated. As a result, GCs occur frequently.

Procedure

Check the GC duration.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** > **ALM-18011 NodeManager GC Time Exceeds the Threshold** > **Location** to check the IP address of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Yarn** > **Instance** > **NodeManager (IP address for which the alarm is generated)**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **Garbage Collection (GC) Time of NodeManager** to check the GC duration statistics of the Broker process collected every minute.

Step 3 Check whether the GC duration of the NodeManager process collected every minute exceeds the threshold (12 seconds by default).

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Yarn** > **Configurations** > **All Configurations** > **NodeManager** > **System** to increase the value of **GC_OPTS** parameter as required.

NOTE

The mapping between the number of NodeManager instances in a cluster and the memory size of NodeManager is as follows:

- If the number of NodeManager instances in the cluster reaches 100, the recommended JVM parameters for NodeManager instances are as follows: -Xms2G -Xmx4G -XX:NewSize=512M -XX:MaxNewSize=1G
- If the number of NodeManager instances in the cluster reaches 200, the recommended JVM parameters for NodeManager instances are as follows: -Xms4G -Xmx4G -XX:NewSize=512M -XX:MaxNewSize=1G
- If the number of NodeManager instances in the cluster reaches 500, the recommended JVM parameters for NodeManager instances are as follows: -Xms8G -Xmx8G -XX:NewSize=1G -XX:MaxNewSize=2G

Step 5 Save the configuration and restart the NodeManager instance.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **NodeManager** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.135 ALM-18012 JobHistoryServer GC Time Exceeds the Threshold

Description

The system checks the garbage collection (GC) duration of the JobHistoryServer process every 60 seconds. This alarm is generated when the GC duration exceeds the threshold (12 seconds by default).

This alarm is cleared when the GC duration is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18012	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

A long GC duration of the JobHistoryServer process may interrupt the services.

Possible Causes

The heap memory of the JobHistoryServer instance is overused or the heap memory is inappropriately allocated. As a result, GCs occur frequently.

Procedure

Check the GC duration.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** > **ALM-18012 JobHistoryServer GC Time Exceeds the Threshold** > **Location** to check the IP address of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **MapReduce** > **Instance** > **JobHistoryServer (IP address for which the alarm is generated)**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **Garbage Collection (GC) Time of the JobHistoryServer** to check the GC duration statistics of the Broker process collected every minute.

Step 3 Check whether the GC duration of the JobHistoryServer process collected every minute exceeds the threshold (12 seconds by default).

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **MapReduce** > **Configurations** > **All Configurations** > **JobHistoryServer** > **System** to increase the value of **GC_OPTS** parameter as required.

NOTE

The mapping between the number of historical tasks (10000) and the memory of the JobHistoryServer is as follows:

-Xms30G -Xmx30G -XX:NewSize=1G -XX:MaxNewSize=2G

Step 5 Save the configuration and restart the JobHistoryServer instance.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **JobHistoryServer** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.136 ALM-18013 ResourceManager Direct Memory Usage Exceeds the Threshold

Description

The system checks the direct memory usage of the Yarn service every 30 seconds. This alarm is generated when the direct memory usage of a ResourceManager instance exceeds the threshold (90% of the maximum memory).

The alarm is cleared when the direct memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18013	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available direct memory of the Yarn service is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The direct memory of the ResourceManager instance is overused or the direct memory is inappropriately allocated.

Procedure

Check the direct memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** > **ALM-18013 ResourceManager Direct Memory Usage Exceeds the Threshold** > **Location** to check the IP address of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Yarn** > **Instance** > **ResourceManager (IP address for which the alarm is generated)**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **Memory Usage Status of ResourceManager** to check the direct memory usage.

Step 3 Check whether the used direct memory of ResourceManager reaches 90% of the maximum direct memory specified for ResourceManager by default.

- If yes, go to **Step 4**.
- If no, go to **Step 9**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Yarn** > **Configurations** > **All Configurations** > **ResourceManager** > **System** to increase the value of check whether **-XX:MaxDirectMemorySize** exists in the **GC_OPTS** parameter.

- If yes, go to **Step 5**.
- If no, go to **Step 7**.

Step 5 In the **GC_OPTS** parameter, delete **-XX:MaxDirectMemorySize**.

Step 6 Save the configuration and restart the ResourceManager instance.

Step 7 Check whether the **ALM-18008 Heap Memory Usage of ResourceManager Exceeds the Threshold** exists.

- If yes, handle the alarm by referring to **ALM-18008 Heap Memory Usage of ResourceManager Exceeds the Threshold**.
- If no, go to **Step 8**.

Step 8 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect fault information.

Step 9 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 10 Select **ResourceManager** in the required cluster from the **Service**.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.137 ALM-18014 NodeManager Direct Memory Usage Exceeds the Threshold

Description

The system checks the direct memory usage of the Yarn service every 30 seconds. This alarm is generated when the direct memory usage of a NodeManager instance exceeds the threshold (90% of the maximum memory).

The alarm is cleared when the direct memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18014	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available direct memory of the Yarn service is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The direct memory of the NodeManager instance is overused or the direct memory is inappropriately allocated.

Procedure

Check the direct memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm > Alarms** > **ALM-18014 NodeManager Direct Memory Usage Exceeds the Threshold** > **Location** to check the IP address of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Yarn** > **Instance** > **NodeManager (IP address for which the alarm is generated)**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **Resource** > **Percentage of Used Memory of the NodeManager** to check the direct memory usage.

Step 3 Check whether the used direct memory of NodeManager reaches 90% of the maximum direct memory specified for NodeManager by default.

- If yes, go to **Step 4**.
- If no, go to **Step 9**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Yarn** > **Configurations** > **All Configurations**> **NodeManager** > **System** to check whether "-XX:MaxDirectMemorySize" exists in the **GC_OPTS** parameter.

- If yes, go to **Step 5**.
- If no, go to **Step 7**.

Step 5 In the **GC_OPTS** parameter, delete "-XX:MaxDirectMemorySize".

Step 6 Save the configuration and restart the NodeManager instance.

Step 7 Check whether the **ALM-18018 NodeManager Heap Memory Usage Exceeds the Threshold** exists.

- If yes, handle the alarm by referring to **ALM-18018 NodeManager Heap Memory Usage Exceeds the Threshold**.
- If no, go to **Step 8**.

Step 8 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect fault information.

Step 9 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 10 Select **NodeManager** in the required cluster from the **Service**.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.138 ALM-18015 JobHistoryServer Direct Memory Usage Exceeds the Threshold

Description

The system checks the direct memory usage of the MapReduce service every 30 seconds. This alarm is generated when the direct memory usage of a JobHistoryServer instance exceeds the threshold (90% of the maximum memory).

The alarm is cleared when the direct memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18015	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available direct memory of the MapReduce service is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The direct memory of the JobHistoryServer instance is overused or the direct memory is inappropriately allocated.

Procedure

Check the direct memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** > **ALM-18015 JobHistoryServer Direct Memory Usage Exceeds the Threshold** > **Location** to check the IP address of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **MapReduce** > **Instance** > **JobHistoryServer (IP address for which the alarm is generated).** Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **Memory Usage Status of JobHistoryServer** to check the direct memory usage.

Step 3 Check whether the used direct memory of JobHistoryServer reaches 90% of the maximum direct memory specified for JobHistoryServer by default.

- If yes, go to **Step 4**.
- If no, go to **Step 9**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **MapReduce** > **Configurations** > **All Configurations** > **JobHistoryServer** > **System** to check whether "-XX:MaxDirectMemorySize" exists in the **GC_OPTS** parameter.

- If yes, go to **Step 5**.
- If no, go to **Step 7**.

Step 5 In the **GC_OPTS** parameter, delete "-XX:MaxDirectMemorySize".

Step 6 Save the configuration and restart the JobHistoryServer instance.

Step 7 Check whether the **ALM-18009 Heap Memory Usage of JobHistoryServer Exceeds the Threshold** exists.

- If yes, handle the alarm by referring to **ALM-18009 Heap Memory Usage of JobHistoryServer Exceeds the Threshold**.
- If no, go to **Step 8**.

Step 8 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect fault information.

Step 9 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 10 Select **JobHistoryServer** in the required cluster from the **Service**.

Step 11 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.139 ALM-18016 Non Heap Memory Usage of ResourceManager Exceeds the Threshold

Description

The system checks the Non Heap memory usage of Yarn ResourceManager every 30 seconds and compares the actual usage with the threshold. The alarm is generated when the Non Heap memory usage of Yarn ResourceManager exceeds the threshold (90% of the maximum memory by default).

Users can choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Yarn** to change the threshold.

The alarm is cleared when the Non Heap memory usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18016	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.

Name	Meaning
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

When the Non Heap memory usage of Yarn ResourceManager is overhigh, the performance of Yarn task submission and operation is affected. In addition, a memory overflow may occur so that the Yarn service is unavailable.

Possible Causes

The Non Heap memory of the Yarn ResourceManager instance on the node is overused or the Non Heap memory is inappropriately allocated. As a result, the usage exceeds the threshold.

Procedure

Check the Non Heap memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** > **ALM-18016 Non Heap Memory Usage of Yarn ResourceManager Exceeds the Threshold** > **Location**. Check the HostName of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Yarn** > **Instance** > **ResourceManager**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **Percentage of Used Memory of the ResourceManager**. ResourceManager indicates the corresponding HostName of the instance for which the alarm is generated. Check the Non Heap memory usage.

Step 3 Check whether the used Non Heap memory of ResourceManager reaches 90% of the maximum Non Heap memory specified for ResourceManager by default.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Yarn** > **Configurations** > **All Configurations** > **ResourceManager** > **System**. Adjust the **GC_OPTS** memory parameter of ResourceManager. Save the configuration and restart the ResourceManager instance.

NOTE

The mapping between the number of NodeManager instances in a cluster and the memory size of ResourceManager is as follows:

- If the number of NodeManager instances in the cluster reaches 100, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms4G -Xmx4G -XX:NewSize=512M -XX:MaxNewSize=1G
- If the number of NodeManager instances in the cluster reaches 200, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms6G -Xmx6G -XX:NewSize=512M -XX:MaxNewSize=1G
- If the number of NodeManager instances in the cluster reaches 500, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms10G -Xmx10G -XX:NewSize=1G -XX:MaxNewSize=2G
- If the number of NodeManager instances in the cluster reaches 1000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms20G -Xmx20G -XX:NewSize=1G -XX:MaxNewSize=2G
- If the number of NodeManager instances in the cluster reaches 2000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms40G -Xmx40G -XX:NewSize=2G -XX:MaxNewSize=4G
- If the number of NodeManager instances in the cluster reaches 3000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms60G -Xmx60G -XX:NewSize=2G -XX:MaxNewSize=4G
- If the number of NodeManager instances in the cluster reaches 4000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms80G -Xmx80G -XX:NewSize=2G -XX:MaxNewSize=4G
- If the number of NodeManager instances in the cluster reaches 5000, the recommended JVM parameters of the ResourceManager instance are as follows: -Xms100G -Xmx100G -XX:NewSize=3G -XX:MaxNewSize=6G

Step 5 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 7 Select the following node in the required cluster from the **Service**.

- NodeAgent
- Yarn

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.140 ALM-18017 Non Heap Memory Usage of NodeManager Exceeds the Threshold

Description

The system checks the Non Heap memory usage of Yarn NodeManager every 30 seconds and compares the actual usage with the threshold. The alarm is generated when the Non Heap memory usage of Yarn NodeManager exceeds the threshold (90% of the maximum memory by default).

Users can choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Yarn** to change the threshold.

The alarm is cleared when the Non Heap memory usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18017	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

When the Non Heap memory usage of Yarn NodeManager is overhigh, the performance of Yarn task submission and operation is affected. In addition, a memory overflow may occur so that the Yarn service is unavailable.

Possible Causes

The Non Heap memory of the Yarn NodeManager instance on the node is overused or the Non Heap memory is inappropriately allocated. As a result, the usage exceeds the threshold.

Procedure

Check the Non Heap memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** > **ALM-18017 Non Heap Memory Usage of Yarn NodeManager Exceeds the Threshold** > **Location**. Check the HostName of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Yarn** > **Instance** > **NodeManager**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **Resource** > **Percentage of Used Memory of the NodeManager**. NodeManager indicates the corresponding HostName of the instance for which the alarm is generated. Check the Non Heap memory usage.

Step 3 Check whether the used Non Heap memory of NodeManager reaches 90% of the maximum Non Heap memory specified for NodeManager by default.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Yarn** > **Configurations** > **All Configurations** > **NodeManager** > **System**. Adjust the **GC_OPTS** memory parameter of NodeManager, click **Save**, and click **OK,** and restart the role instance.

NOTE

The mapping between the number of NodeManager instances in a cluster and the memory size of NodeManager is as follows:

- If the number of NodeManager instances in the cluster reaches 100, the recommended JVM parameters for NodeManager instances are as follows: -Xms2G -Xmx4G -XX:NewSize=512M -XX:MaxNewSize=1G
- If the number of NodeManager instances in the cluster reaches 200, the recommended JVM parameters for NodeManager instances are as follows: -Xms4G -Xmx4G -XX:NewSize=512M -XX:MaxNewSize=1G
- If the number of NodeManager instances in the cluster reaches 500, the recommended JVM parameters for NodeManager instances are as follows: -Xms8G -Xmx8G -XX:NewSize=1G -XX:MaxNewSize=2G

Step 5 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 7 Select the following node in the required cluster from the **Service**.

- NodeAgent
- Yarn

Step 8 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.141 ALM-18018 NodeManager Heap Memory Usage Exceeds the Threshold

Description

The system checks the heap memory usage of Yarn NodeManager every 30 seconds and compares the actual usage with the threshold. The alarm is generated when the heap memory usage of Yarn NodeManager exceeds the threshold (95% of the maximum memory by default).

The alarm is cleared when the heap memory usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18018	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.

Name	Meaning
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

When the heap memory usage of Yarn NodeManager is overhigh, the performance of Yarn task submission and operation is affected. In addition, a memory overflow may occur so that the Yarn service is unavailable.

Possible Causes

The heap memory of the Yarn NodeManager instance on the node is overused or the heap memory is inappropriately allocated. As a result, the usage exceeds the threshold.

Procedure

Check the heap memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** > **ALM-18018 NodeManager Heap Memory Usage Exceeds the Threshold** > **Location**. Check the HostName of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Yarn** > **Instance** > **NodeManager**. Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **Resource** > **Percentage of Used Memory of the NodeManager** to check the heap memory usage.

Step 3 Check whether the used heap memory of NodeManager reaches 95% of the maximum heap memory specified for NodeManager.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Yarn** > **Configurations** > **All Configurations** > **NodeManager** > **System**. Increase the value of **GC_OPTS** parameter as required, click **Save**, and click **OK**, and restart the role instance.

NOTE

The mapping between the number of NodeManager instances in a cluster and the memory size of NodeManager is as follows:

- If the number of NodeManager instances in the cluster reaches 100, the recommended JVM parameters for NodeManager instances are as follows: -Xms2G -Xmx4G -XX:NewSize=512M -XX:MaxNewSize=1G
- If the number of NodeManager instances in the cluster reaches 200, the recommended JVM parameters for NodeManager instances are as follows: -Xms4G -Xmx4G -XX:NewSize=512M -XX:MaxNewSize=1G
- If the number of NodeManager instances in the cluster reaches 500, the recommended JVM parameters for NodeManager instances are as follows: -Xms8G -Xmx8G -XX:NewSize=1G -XX:MaxNewSize=2G

Step 5 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 7 Select the following node in the required cluster from the **Service**.

- NodeAgent
- Yarn

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.142 ALM-18019 Non Heap Memory Usage of JobHistoryServer Exceeds the Threshold

Description

The system checks the Non Heap memory usage of MapReduce JobHistoryServer every 30 seconds and compares the actual usage with the threshold. The alarm is generated when the Non Heap memory usage of MapReduce JobHistoryServer exceeds the threshold (90% of the maximum memory by default).

Users can choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **MapReduce** to change the threshold.

The alarm is cleared when the Non Heap memory usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18019	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

When the Non Heap memory usage of MapReduce JobHistoryServer is overhigh, the performance of MapReduce task submission and operation is affected. In addition, a memory overflow may occur so that the MapReduce service is unavailable.

Possible Causes

The Non Heap memory of the MapReduce JobHistoryServer instance on the node is overused or the Non Heap memory is inappropriately allocated. As a result, the usage exceeds the threshold.

Procedure

Check the Non Heap memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** > **ALM-18019 Non Heap Memory Usage of MapReduce JobHistoryServer Exceeds the Threshold** > **Location**. Check the HostName of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **MapReduce** > **Instance** > **JobHistoryServer.** Click the drop-down menu in the upper right corner of **Chart**, choose **Customize** > **JobHistoryServer Non Heap memory usage statistics**. JobHistoryServer indicates the corresponding HostName of the instance for which the alarm is generated. Check the Non Heap memory usage.

Step 3 Check whether the used Non Heap memory of JobHistoryServer reaches 90% of the maximum Non Heap memory specified for JobHistoryServer.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **MapReduce** > **Configurations** > **All Configurations** > **JobHistoryServer** > **System**. Adjust the **GC_OPTS** memory parameter of the NodeManager, click **Save**, and click **OK,** and restart the role instance.

NOTE

The mapping between the number of historical tasks (10000) and the memory of the JobHistoryServer is as follows:

-Xms30G -Xmx30G -XX:NewSize=1G -XX:MaxNewSize=2G

Step 5 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 7 Select the following node in the required cluster from the **Service**.

- NodeAgent
- MapReduce

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.143 ALM-18020 Yarn Task Execution Timeout

Description

The system checks MapReduce and Spark tasks (except for permanent JDBC tasks) submitted to Yarn every 15 minutes. This alarm is generated when the task execution time exceeds the timeout duration specified by the user. However, the task can be properly executed. The client timeout parameter of MapReduce is mapreduce.application.timeout.alarm and that of Spark is spark.application.timeout.alarm. The unit is ms.

This alarm is cleared when the task is finished or terminated.

Attribute

Alarm ID	Alarm Severity	Auto Clear
18020	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
ApplicationName	Specifies the object (application ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

The alarm persists after task execution times out. However, the task can still be properly executed, so this alarm does not exert any impact on the system.

Possible Causes

- The specified timeout duration is shorter than the required execution time.
- The queue resources for task running are insufficient.
- Task data skew occurs. As a result, some tasks process a large amount of data and take a long time to execute.

Procedure

Check whether the timeout interval is correctly set.

Step 1 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Alarm** > **Alarms**. The **Alarms** page is displayed.

Step 2 Select the alarm whose ID is **18020**. In the alarm details, view **Location** to obtain the timeout task name and timeout duration.

Step 3 Based on the task name and timeout interval, choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn** > **ResourceManager (Active)** to log in to the native Yarn page. Then find the task on the native page, check its **StartTime** and calculate the task execution time based on the current system time. Check whether the task execution time exceeds the timeout duration.

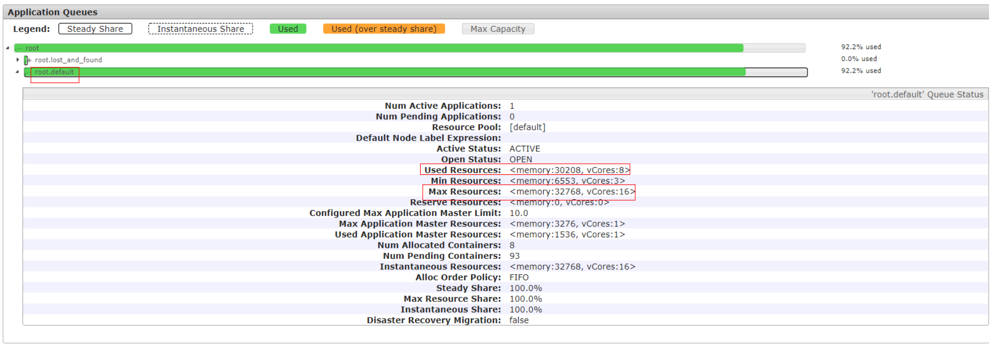
- If yes, go to **Step 5**.
- If no, go to **Step 10**.

Step 4 Evaluate the expected task execution time based on the service and compare it with the task timeout interval. If the timeout interval is too short, set the timeout interval (**mapreduce.application.timeout.alarm** or **spark.application.timeout.alarm**) of the client to the task expected execution time. Run the task again and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check whether the queue resources are sufficient.

Step 5 Find the task on the native page and view the queue name of the task. Click **Scheduler** on the left of the native page. On the **Applications Queues** page, find the corresponding queue name and expand the queue details, as shown in the following figure.



'root.default' Queue Status

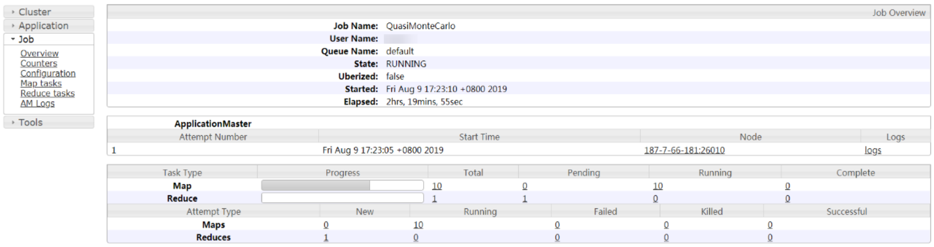
Num Active Applications: 1
Num Pending Applications: 0
Resource Pool: [default]
Default Node Label Expression:
Active Status: ACTIVE
Open Status: OPEN
Used Resources: <memory:30208, vCores:8>
Min Resources: <memory:6553, vCores:3>
Max Resources: <memory:32768, vCores:16>
Reserve Resources: <memory:0, vCores:0>
Configured Max Application Master Limit: 10.0
Max Application Master Resources: <memory:3276, vCores:1>
Used Application Master Resources: <memory:1536, vCores:1>
Num Allocated Containers: 8
Num Pending Containers: 93
Instantaneous Resources: <memory:32768, vCores:16>
Alloc Order Policy: FIFO
Steady Share: 100.0%
Max Resource Share: 100.0%
Instantaneous Share: 100.0%
Disaster Recovery Migration: false

Step 6 Check whether the value of **Used Resources** in the queue details is approximately equal to the value of **Max Resources**, which indicates that the resources in the queue submitted by the task have been used up. If the queue resources are insufficient, choose **Tenant Resources** > **Dynamic Resource Plan** > **Resource Distribution Policy** on FusionInsight Manager and increase the value of **Max Resources** for the queue. Run the task again and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Check whether data skew occurs.

Step 7 On the native Yarn page, click *task ID* (for example, **application_1565337919723_0002**) > **Tracking URL:ApplicationMaster** > **job_1565337919723_0002**. The following page is displayed.



Job Name:	QuasiMonteCarlo
User Name:	[redacted]
Queue Name:	default
State:	RUNNING
Uberized:	false
Started:	Fri Aug 9 17:23:10 +0800 2019
Elapsed:	2hrs, 19mins, 55sec

ApplicationMaster

Attempt Number	Start Time	Node	Logs
1	Fri Aug 9 17:23:05 +0800 2019	187-7-66-181:26010	logs

Task Type	Progress	Total	Pending	Running	Complete
Map		10	0	10	0
Reduce		1	1	0	0

Attempt Type	New	Running	Failed	Killed	Successful
Maps	0	10	0	0	0
Reduces	1	0	0	0	0

Step 8 Choose **Job** > **Map tasks** or **Job** > **Reduce tasks** on the left and check whether the execution time of each Map or Reduce task differs greatly. If yes, task data skew occurs. In this case, you need to balance the task data.

Step 9 Rectify the fault based on the preceding causes and perform the tasks again. Then, check whether the alarm persists.

- If yes, go to **Step 10**.
- If no, no further action is required.

Collect the fault information.

Step 10 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 11 Expand the **Service** drop-down list, and select **Yarn** for the target cluster.

Step 12 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 13 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.144 ALM-18021 Mapreduce Service Unavailable

Description

The alarm module checks the MapReduce service status every 60 seconds. This alarm is generated when the system detects that the MapReduce service is unavailable.

The alarm is cleared when the MapReduce service recovers.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
18021	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The cluster cannot provide the MapReduce service. For example, MapReduce cannot be used to view task logs or the log archive function is unavailable.

Possible Causes

- The JobHistoryServer instance is abnormal.
- The KrbServer service is abnormal.
- The ZooKeeper service abnormal.
- The HDFS service abnormal.
- The Yarn service is abnormal.

Procedure

Check MapReduce service JobHistoryServer instance status.

Step 1 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **MapReduce** > **Instance**.

Step 2 Check whether the running status of JobHistoryServer is **Normal**.

- If yes, go to **Step 11**.
- If no, go to **Step 3**.

Check the KrbServer service status.

Step 3 In the alarm list on FusionInsight Manager, check whether **ALM-25500 KrbServer Service Unavailable** exists.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 Rectify the fault by following the steps provided in **ALM-25500 KrbServer Service Unavailable**, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check the ZooKeeper service.

Step 5 In the alarm list on FusionInsight Manager, check whether **ALM-13000 ZooKeeper Service Unavailable** exists.

- If yes, go to **Step 6**.
- If no, go to **Step 7**.

Step 6 Rectify the fault by following the steps provided in **ALM-13000 ZooKeeper Service Unavailable**, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Check the HDFS service status.

Step 7 In the alarm list on FusionInsight Manager, check whether **ALM-14000 HDFS Service Unavailable** exists.

- If yes, go to **Step 8**.
- If no, go to **Step 9**.

Step 8 Rectify the fault by following the steps provided in **ALM-14000 HDFS Service Unavailable**, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Check the Yarn service status.

Step 9 In the alarm list on FusionInsight Manager, check whether **ALM-18000 Yarn Service Unavailable** exists.

- If yes, go to **Step 10**
- If no, go to **Step 11**.

Step 10 Rectify the fault by following the steps provided in **ALM-18000 Yarn Service Unavailable**, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Collect fault information.

Step 11 On the FusionInsight Manager home page of the active cluster, choose **O&M** > **Log** > **Download.**

Step 12 Select **MapReduce** in the required cluster from the **Service.**

Step 13 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 14 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.145 ALM-18022 Insufficient Yarn Queue Resources

Description

The alarm module checks Yarn queue resources every 60 seconds. This alarm is generated when available resources or ApplicationMaster (AM) resources of a queue are insufficient.

This alarm is cleared when available resources are sufficient.

Attribute

Alarm ID	Alarm Severity	Auto Clear
18022	Minor	Yes

Parameters

Parameter Name	Description
Source	Specifies the cluster for which the alarm is generated.
QueueName	Specifies the queue for which the alarm is generated.
QueueMetric	Specifies the metric of the queue for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

- An application being executed takes longer time.
- An application fails to be executed for a long time after being submitted.

Possible Causes

- NodeManager node resources are insufficient.
- The configured maximum resource capacity of the queue is excessively small.
- The configured maximum AM resource percentage is excessively small.

Procedure

View alarm details.

Step 1 On the FusionInsight Manager, choose **O&M** > **Alarm > Alarms**.

Step 2 View location information of this alarm and check whether **QueueName** is **root** and **QueueMetric** is **Memory** or **QueueName** is **root** and **QueueMetric** is **vCores**.

- If yes, go to **Step 3**.
- If no, go to **Step 4**.

Step 3 The memory or CPU of the Yarn cluster is insufficient. In this case, log in to the node where NodeManager resides and run the **free -g** and **cat /proc/cpuinfo** commands to query the available memory and available CPU of the node, respectively. On FusionInsight Manager, increase the values of **yarn.nodemanager.resource.memory-mb** and **yarn.nodemanager.resource.cpu-vcores** for the Yarn NodeManager based on the query results. Then, restart the NodeManager instance. Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Step 4 View location information of this alarm and check whether **QueueName** is **<*Tenant Queue*>** and **QueueMetric** is **Memory**, or **QueueName** is **<*Tenant Queue*>** and **QueueMetric** is **vCores** in **Location**, check whether **available Memory =** or **available vCores =** are included in **Additional Information**.

- If yes, go to **Step 5**.
- If no, go to **Step 7**.

Step 5 The memory or CPU of the tenant queue is insufficient. In this case, choose **Tenant Resources** > **Dynamic Resource Plan > Resource Distribution Policy** and increase the value of **Maximum Capacity**. Then, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Step 6 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn** > **Configurations** > **All Configurations**. Enter the keyword "threshold" and click **ResourceManager**. Adjust the threshold values of the following parameters:

If **Additional Information** contains **available Memory =**, change the value of **yarn.queue.memory.alarm.threshold** to a value smaller than that of **available Memory =** in **Additional Information**.

If **Additional Information** contains **available vCores =**, change the value of **yarn.queue.vcore.alarm.threshold** to a value smaller than that of **available vCores =** in **Additional Information**.

Wait for five minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Step 7 If **available AmMemory =** or **available AmvCores =** is included in **Additional Information**, ApplicationMaster memory or CPU of the tenant queue is insufficient. In this case, choose **Tenant Resources** > **Dynamic Resource Plan** > **Queue Configuration** and increase the value of **Maximum Am Resource Percent**. Then, check whether this alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Step 8 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn** > **Configurations** > **All Configurations**. Enter the keyword "threshold" and click **ResourceManager**. Adjust the threshold values of the following parameters:

If **Additional Information** contains **available AmMemory =**, change the value of **yarn.queue.memory.alarm.threshold** to a value smaller than that of **available AmMemory =** in **Additional Information**.

If **Additional Information** contains **available AmvCores =**, change the value of **yarn.queue.vcore.alarm.threshold** to a value smaller than that of **available AmvCores =** in **Additional Information**.

Wait for five minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect fault information.

Step 9 Log in to FusionInsight Manager of the active cluster, and choose **O&M** > **Log** > **Download**.

Step 10 Select **Yarn** in the required cluster from the **Service**.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Reference

None

9.13.146 ALM-18023 Number of Pending Yarn Tasks Exceeds the Threshold

Description

The alarm module checks the number of pending applications in the Yarn root queue every 60 seconds. The alarm is generated when the number exceeds 60.

Attribute

Alarm ID	Alarm Severity	Auto Clear
18023	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
QueueName	Identifies the queue for which the alarm is generated.
QueueMetric	Identifies the queue indicator for which the alarm is generated.

Impact on the System

- It takes long time to end an application.
- A new application cannot run after submission.

Possible Causes

- NodeManager node resources are insufficient.
- The maximum resource capacity of the queue and the maximum AM resource percentage are too small.
- The monitoring threshold is too small.

Procedure

Check NodeManager resources.

Step 1 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn** > **ResourceManager (Active)** to access the ResourceManager web UI.

Step 2 Click **Scheduler** and check whether the root queue resources are used up in **Application Queues**.

- If yes, go to **Step 3**.
- If no, go to **Step 4**.

Step 3 Expand the capacity of the NodeManager instance of the Yarn service. After the capacity expansion, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Check the maximum queue resource capacity and the maximum AM resource percentage.

Step 4 Check whether the resources of the queue corresponding to the pending task are used up.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 On FusionInsight Manager, choose **Tenant Resources** > **Dynamic Resource Plan** and add resources as required. Check whether the alarms are cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Adjust the monitoring thresholds.

Step 6 On FusionInsight Manager, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Yarn** > **Applications** > **Pending Applications**, and increase the thresholds as required.

Step 7 Check whether the alarm is cleared 5 minutes later.

- If yes, no further action is required.
- If no, go to **Step 8**.

Collect the fault information.

Step 8 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 9 Expand the **Service** drop-down list, and select **Yarn** for the target cluster.

Step 10 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.147 ALM-18024 Pending Yarn Memory Usage Exceeds the Threshold

Description

The alarm module checks the pending memory of Yarn every 60 seconds. The alarm is generated when the pending memory exceeds the threshold. Pending memory indicates the total memory that is not allocated to submitted Yarn applications.

Attribute

Alarm ID	Alarm Severity	Auto Clear
18024	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
QueueName	Identifies the queue for which the alarm is generated.
QueueMetric	Identifies the queue indicator for which the alarm is generated.

Impact on the System

- It takes long time to end an application.
- A new application cannot run after submission.

Possible Causes

- NodeManager node resources are insufficient.
- The maximum resource capacity of the queue and the maximum AM resource percentage are too small.
- The monitoring threshold is too small.

Procedure

Check NodeManager resources.

Step 1 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn** > **ResourceManager (Active)** to access the ResourceManager web UI.

Step 2 Click **Scheduler** and check whether the root queue resources are used up in **Application Queues**.

- If yes, go to **Step 3**.
- If no, go to **Step 4**.

Step 3 Expand the capacity of the NodeManager instance of the Yarn service. After the capacity expansion, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Check the maximum queue resource capacity and the maximum AM resource percentage.

Step 4 Check whether the resources of the queue corresponding to the pending task are used up.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 On FusionInsight Manager, choose **Tenant Resources** > **Dynamic Resource Plan** and add resources as required. Check whether the alarms are cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Adjust the monitoring thresholds.

Step 6 On FusionInsight Manager, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Yarn** > **CPU and Memory** > **Pending Memory**, and increase the threshold as required.

Step 7 Check whether the alarm is cleared 5 minutes later.

- If yes, no further action is required.
- If no, go to **Step 8**.

Collect the fault information.

Step 8 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 9 Expand the **Service** drop-down list, and select **Yarn** for the target cluster.

Step 10 Click in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.148 ALM-18025 Number of Terminated Yarn Tasks Exceeds the Threshold

Description

The alarm module checks the number of terminated applications in the Yarn root queue every 60 seconds. The alarm is generated when the number exceeds 50 for three consecutive times.

Attribute

Alarm ID	Alarm Severity	Auto Clear
18025	Major	Yes

Parameters

Name	Meaning
Cluster Name	Specifies the cluster for which the alarm is generated.
Service Name	Specifies the service for which the alarm is generated.
Role Name	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

A large number of application tasks are forcibly terminated.

Possible Causes

- The user forcibly terminates a large number of tasks.
- The system terminates tasks due to some error.

Procedure

Check the alarm details.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** to go to the alarm page.

Step 2 View **Additional Information** in the alarm details to check whether the alarm threshold is too small.

- If yes, go to **Step 3**.
- If no, go to **Step 4**.

Step 3 Choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Yarn** > **Other** > **Terminated Applications of root queue** to modify the threshold. Go to **Step 6**.

Step 4 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn** > **ResourceManager(Active)** to access the ResourceManager web UI.

Step 5 Click **KILLED** in **Applications** and click the task on the top. View the description of **Diagnostics** and rectify the fault based on the task termination details (for example, the task is terminated by a user).

Step 6 Wait for 3 minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect the fault information.

Step 7 On the FusionInsight Manager, choose **O&M > Log > Download**.

Step 8 Expand the **Service** drop-down list, and select **Yarn** for the target cluster.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.149 ALM-18026 Number of Failed Yarn Tasks Exceeds the Threshold

Description

The alarm module checks the number of failed applications in the Yarn root queue every 60 seconds. The alarm is generated when the number exceeds 50 for three consecutive times.

Attribute

Alarm ID	Alarm Severity	Auto Clear
18026	Major	Yes

Parameters

Name	Meaning
Cluster Name	Specifies the cluster for which the alarm is generated.
Service Name	Specifies the service for which the alarm is generated.
Role Name	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

- A large number of application tasks fail to be executed.
- Failed tasks need to be submitted again.

Possible Causes

The task fails to be executed due to some error.

Procedure

Check the alarm details.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** to go to the alarm page.

Step 2 View **Additional Information** in the alarm details to check whether the alarm threshold is too small.

- If yes, go to **Step 3**.
- If no, go to **Step 4**.

Step 3 Choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Yarn** > **Other** > **Failed Applications of root queue** to modify the threshold. Go to **Step 6**.

Step 4 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn** > **ResourceManager(Active)** to access the ResourceManager web UI.

Step 5 Click **FAILED** in **Applications** and click the task on the top. View the description of **Diagnostics** and rectify the fault based on the task failure causes.

Step 6 Wait for 3 minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect the fault information.

Step 7 On the FusionInsight Manager, choose O&M > Log > Download.

Step 8 Expand the **Service** drop-down list, and select **Yarn** for the target cluster.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.150 ALM-19000 HBase Service Unavailable

Description

This alarm is generated when the HBase service is unavailable. The alarm module checks the HBase service status every 120 seconds.

This alarm is cleared when the HBase service recovers.

NOTE

If the multi-instance function is enabled in the cluster and multiple HBase service instances are installed, you need to determine the HBase service instance where the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the HBase1 service is unavailable, ServiceName=HBase1 is displayed in **Location**, and the operation object in the procedure needs to be changed from HBase to HBase1.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
19000	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

Operations, such as reading or writing data and creating tables, cannot be performed.

Possible Causes

- The ZooKeeper service is abnormal.
- The HDFS service is abnormal.
- The HBase service is abnormal.
- The network is abnormal.

Procedure

Check the ZooKeeper service status.

Step 1 On the FusionInsight Manager, check whether the running status of ZooKeeper is **Normal** on service list.

- If yes, go to **Step 5**.
- If no, go to **Step 2**.

Step 2 In the alarm list, check whether **ALM-13000 ZooKeeper Service Unavailable** exists.

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 Rectify the fault by following the steps provided in **ALM-13000 ZooKeeper Service Unavailable**.

Step 4 Wait several minutes, and check whether alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check the HDFS service status.

Step 5 In the alarm list, check whether **ALM-14000 HDFS Service Unavailable** exists.

- If yes, go to **Step 6**.
- If no, go to **Step 8**.

Step 6 Rectify the fault by following the steps provided in **ALM-14000 HDFS Service Unavailable**.

Step 7 Wait several minutes, and check whether alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Step 8 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS**. Check whether **Safe Mode** is **ON**.

- If yes, go to **Step 9**.
- If no, go to **Step 12**.

Step 9 Log in to the HDFS client as user **root**. Run **cd** to switch to the client installation directory, and run **source bigdata_env**.

If the cluster uses the security mode, perform security authentication. Obtain the password of user hdfs from the administrator, run the **kinit hdfs** command and enter the password as prompted.

Step 10 Run the following command to manually exit the safe mode:

```
hdfs dfsadmin -safemode leave
```

Step 11 Wait several minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 12**.

Check the HBase service status.

Step 12 On the FusionInsight Manager portal, click **Cluster** > *Name of the desired cluster* > **Services** > **HBase**.

Step 13 Check whether there is one active HMaster and one standby HMaster.

- If yes, go to **Step 15**.
- If no, go to **Step 14**.

Step 14 Click **Instances**, select the HMaster whose status is not **Active**, click **More**, and select **Restart Instance** to restart the HMaster. Check whether there is one active HMaster and one standby HMaster again.

- If yes, go to **Step 15**.
- If no, go to **Step 21**.

Step 15 Choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase** > **HMaster(Active)** to go to the HMaster WebUI.

NOTE

By default, the **admin** user does not have the permissions to manage other components. If the page cannot be opened or the displayed content is incomplete when you access the native UI of a component due to insufficient permissions, you can manually create a user with the permissions to manage that component.

Step 16 Check whether at least one RegionServer exists under **Region Servers**.

- If yes, go to **Step 17**.
- If no, go to **Step 21**.

Step 17 Check **Tables** > **System Tables**, as shown in **Figure 9-26**. Check whether **hbase:meta**, **hbase:namespace**, and **hbase:acl** exist in the **Table Name** column.

- If yes, go to **Step 18**.
- If no, go to **Step 19**.

Figure 9-26 HBase system table

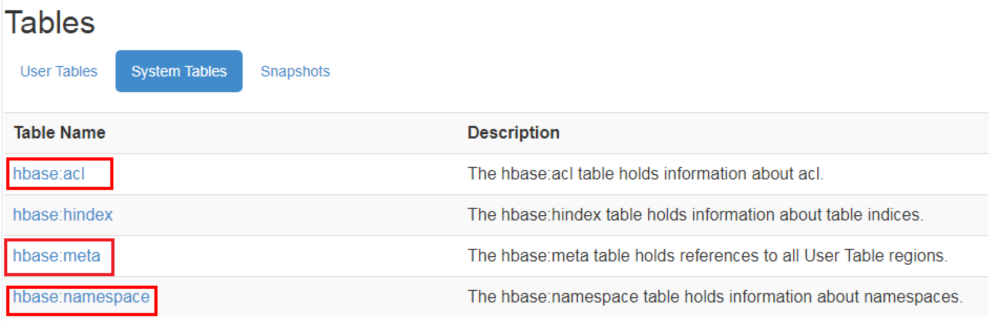


Table Name	Description
hbase:acl	The hbase:acl table holds information about acl.
hbase:hindex	The hbase:hindex table holds information about table indices.
hbase:meta	The hbase:meta table holds references to all User Table regions.
hbase:namespace	The hbase:namespace table holds information about namespaces.

Step 18 As shown in **Figure 9-26**, click the **hbase:meta**, **hbase:namespace**, and **hbase:acl** hyperlinks and check whether the pages are properly displayed. If the pages are properly displayed, the tables are normal.

If they are, go to **Step 19**.

If they are not, go to **Step 23**.

NOTE

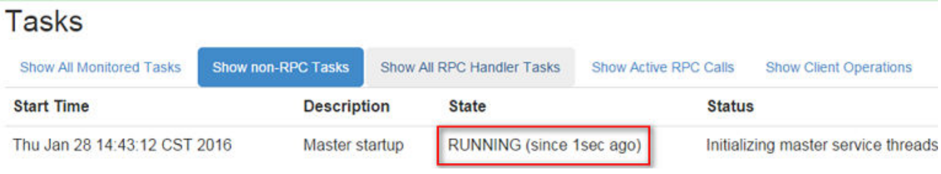
In normal mode, **ACL** is enabled for HBase by default. The **hbase:acl** table is generated only when **ACL** is manually enabled. In this case, check this table. In other scenarios, this table does not need to be checked.

Step 19 View the HMaster startup status.

In **Figure 9-27**, if the **RUNNING** state exists in **Tasks**, HMaster is being started. In the **State** column, you can view the time when HMaster is in the **RUNNING** state. In **Figure 9-28**, if the state is **COMPLETE**, HMaster is started.

Check whether HMaster is in the **RUNNING** state for a long time.

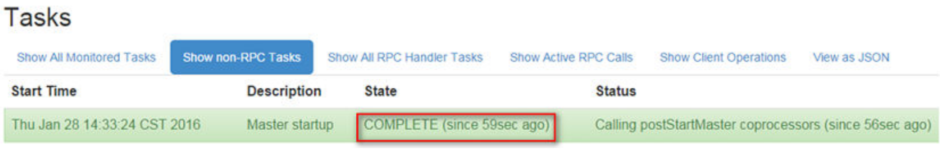
Figure 9-27 HMaster is being started



Tasks

Start Time	Description	State	Status
Thu Jan 28 14:43:12 CST 2016	Master startup	RUNNING (since 1sec ago)	Initializing master service threads

Figure 9-28 HMaster is started



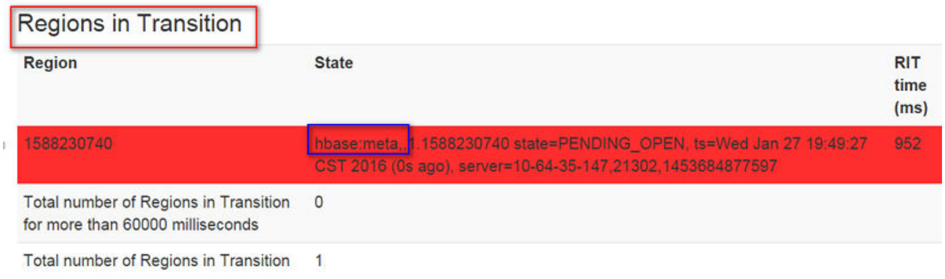
Tasks

Start Time	Description	State	Status
Thu Jan 28 14:33:24 CST 2016	Master startup	COMPLETE (since 59sec ago)	Calling postStartMaster coprocessors (since 56sec ago)

- If yes, go to **Step 20**.
- If no, go to **Step 21**.

Step 20 On the HMaster WebUI, check whether any hbase:meta is in the **Region in Transition** state for a long time.

Figure 9-29 Region in Transition



Regions in Transition

Region	State	RIT time (ms)
1588230740	hbase:meta,,1.1588230740 state=PENDING_OPEN, ts=Wed Jan 27 19:49:27 CST 2016 (0s ago), server=10-64-35-147,21302,1453684877597	952
Total number of Regions in Transition for more than 60000 milliseconds	0	
Total number of Regions in Transition	1	

- If yes, go to **Step 21**.
- If no, go to **Step 22**.

Step 21 In the precondition that services are not affected, log in to the FusionInsight Manager portal and choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase** > **More** > **Restart Service**. Enter the administrator password and click **OK**.

- If yes, go to **Step 22**.
- If no, go to **Step 23**.

Step 22 Wait several minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 23**.

Check the network connection between HMaster and dependent components.

Step 23 On the FusionInsight Manager, choose **Cluster** >*Name of the desired cluster* > **Services** > **HBase**.

Step 24 Click **Instance** and the HMaster instance list is displayed. Record the **management IP Address** in the row of **HMaster(Active)**.

Step 25 Use the IP address obtained in **Step 24** to log in to the host where the active HMaster runs as user **omm** .

Step 26 Run the **ping** command to check whether communication between the host that runs the active HMaster and the hosts that run the dependent components. (The dependent components include ZooKeeper, HDFS and Yarn. Obtain the IP addresses of the hosts that run these services in the same way as that for obtaining the IP address of the active HMaster.)

- If yes, go to **Step 29**.
- If no, go to **Step 27**.

Step 27 Contact the administrator to restore the network.

Step 28 In the alarm list, check whether **HBase Service Unavailable** is cleared.

- If yes, no further action is required.
- If no, go to **Step 29**.

Collect fault information.

Step 29 On the FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 30 Select the following nodes in the required cluster from the **Service** drop-down list:

- ZooKeeper
- HDFS
- HBase

Step 31 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 32 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.151 ALM-19006 HBase Replication Sync Failed

Description

The alarm module checks the HBase DR data synchronization status every 30 seconds. When disaster recovery (DR) data fails to be synchronized to a standby cluster, the alarm is triggered.

When DR data synchronization succeeds, the alarm is cleared.

NOTE

If the multi-instance function is enabled in the cluster and multiple HBase service instances are installed, you need to determine the HBase service instance where the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the HBase1 service is unavailable, **ServiceName=HBase1** is displayed in **Location**, and the operation object in the procedure needs to be changed from HBase to HBase1.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
19006	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

HBase data in a cluster fails to be synchronized to the standby cluster, causing data inconsistency between active and standby clusters.

Possible Causes

- The HBase service on the standby cluster is abnormal.
- A network exception occurs.

Procedure

Observe whether the system automatically clears the alarm.

Step 1 On the FusionInsight Manager portal of the active cluster, click **O&M** > **Alarm** > **Alarms.**

Step 2 In the alarm list, click the alarm to obtain alarm generation time from **Generated** of the alarm. Check whether the alarm has existed for five minutes.

- If yes, go to **Step 4**.
- If no, go to **Step 3**.

Step 3 Wait five minutes and check whether the system automatically clears the alarm.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check the HBase service status of the standby cluster.

Step 4 Log in to the FusionInsight Manager portal of the active cluster, and click **O&M** > **Alarm** > **Alarms.**

Step 5 In the alarm list, click the alarm to obtain **HostName** from **Location**.

Step 6 Access the node where the HBase client of the active cluster resides as user **omm**.

If the cluster uses a security mode, perform security authentication first and then access the **hbase shell** interface as user **hbase**.

cd /opt/client

source ./bigdata_env

kinit *hbaseuser*

Step 7 Run the **status 'replication', 'source'** command to check the DR synchronization status of the faulty node.

The DR synchronization status of a node is as follows.

```
10-10-10-153:
SOURCE: PeerID=abc, SizeOfLogQueue=0, ShippedBatches=2, ShippedOps=2, ShippedBytes=320,
LogReadInBytes=1636, LogEditsRead=5, LogEditsFiltered=3, SizeOfLogToReplicate=0,
TimeForLogToReplicate=0, ShippedHFiles=0, SizeOfHFileRefsQueue=0, AgeOfLastShippedOp=0,
TimeStampsOfLastShippedOp=Mon Jul 18 09:53:28 CST 2016, Replication Lag=0,
FailedReplicationAttempts=0
SOURCE: PeerID=abc1, SizeOfLogQueue=0, ShippedBatches=1, ShippedOps=1, ShippedBytes=160,
LogReadInBytes=1636, LogEditsRead=5, LogEditsFiltered=3, SizeOfLogToReplicate=0,
TimeForLogToReplicate=0, ShippedHFiles=0, SizeOfHFileRefsQueue=0, AgeOfLastShippedOp=16788,
TimeStampsOfLastShippedOp=Sat Jul 16 13:19:00 CST 2016, Replication Lag=16788,
FailedReplicationAttempts=5
```

Step 8 Obtain **PeerID** corresponding to a record whose **FailedReplicationAttempts** value is greater than 0.

In the preceding step, data on the faulty node 10-10-10-153 fails to be synchronized to a standby cluster whose **PeerID** is **abc1**.

Step 9 Run the **list_peers** command to find the cluster and the HBase instance corresponding to the **PeerID** value.

```
PEER_ID CLUSTER_KEY STATE TABLE_CFS
abc1 10.10.10.110,10.10.10.119,10.10.10.133:2181:/hbase2 ENABLED
abc 10.10.10.110,10.10.10.119,10.10.10.133:2181:/hbase ENABLED
```

In the preceding information, **/hbase2** indicates that data is synchronized to the HBase2 instance of the standby cluster.

Step 10 In the service list of FusionInsight Manager of the standby cluster, check whether the running status of the HBase instance obtained by using **Step 9** is **Normal**.

- If yes, go to **Step 14**.
- If no, go to **Step 11**.

Step 11 In the alarm list, check whether the **ALM-19000 HBase Service Unavailable** alarm is generated.

- If yes, go to **Step 12**.
- If no, go to **Step 14**.

Step 12 Follow troubleshooting procedures in **ALM-19000 HBase Service Unavailable** to rectify the fault.

Step 13 Wait for a few minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 14**.

Check network connections between RegionServers on active and standby clusters.

Step 14 Log in to the FusionInsight Manager portal of the active cluster, and click **O&M** > **Alarm** > **Alarms.**

Step 15 In the alarm list, click the alarm to obtain **HostName** from **Location**.

Step 16 Use the IP address obtained in **Step 15** to log in to a faulty RegionServer node as user **omm**.

Step 17 Run the **ping** command to check whether network connections between the faulty RegionServer node and the host where RegionServer of the standby cluster resides are in the normal state.

- If yes, go to **Step 20**.
- If no, go to **Step 18**.

Step 18 Contact the network administrator to restore the network.

Step 19 After the network is running properly, check whether the alarm is cleared in the alarm list.

- If yes, no further action is required.
- If no, go to **Step 20**.

Collect fault information.

Step 20 On the FusionInsight Manager interface of active and standby clusters, choose **O&M** > **Log** > **Download**.

Step 21 In the **Service** drop-down list box, select **HBase** in the required cluster.

Step 22 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 23 Contact the O&M personnel and send the collected fault logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.152 ALM-19007 HBase GC Time Exceeds the Threshold

Description

The system checks the old generation garbage collection (GC) time of the HBase service every 60 seconds. This alarm is generated when the detected old generation GC time exceeds the threshold (exceeds 5 seconds for three consecutive checks by default). To change the threshold, on the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **HBase > GC > GC time for old generation**. This alarm is cleared when the old generation GC time of the HBase service is shorter than or equal to the threshold.

NOTE

If the multi-instance function is enabled in the cluster and multiple HBase service instances are installed, you need to determine the HBase service instance where the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the HBase1 service is unavailable, **ServiceName=HBase1** is displayed in **Location**, and the operation object in the procedure needs to be changed from HBase to HBase1.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
19007	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.

Impact on the System

If the old generation GC time exceeds the threshold, HBase data read and write are affected.

Possible Causes

The memory of HBase instances is overused, the heap memory is inappropriately allocated, or a large number of I/O operations exist in HBase. As a result, GCs occur frequently.

Procedure

Check the GC time.

Step 1 On the FusionInsight Manager portal, click **O&M** > **Alarm** > **Alarms** and select the alarm whose **ID** is **19007**. Then check the role name in **Location** and confirm the IP adress of the instance.

- If the role for which the alarm is generated is HMaster, go to **Step 2**.
- If the role for which the alarm is generated is RegionServer, go to **Step 3**.

Step 2 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase** > **Instance** and click the HMaster for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **GC** > **Garbage Collection (GC) Time of HMaster** and click **OK** to check whether the value of **GC time for old generation** is greater than the threshold (exceeds 5 seconds for three consecutive checks periods by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 3 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase** > **Instance** and click the RegionServer for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **GC** > **Garbage Collection (GC) Time of RegionServer** and click **OK** to check whether the value of **GC time for old generation** is greater than the threshold (exceeds 5 seconds for three consecutive checks periods by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Check the current JVM configuration.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase** > **Configurations**, and click **All Configurations**. In Search, enter **GC_OPTS** to check the **GC_OPTS** memory parameter of role HMaster(HBase->HMaster), RegionServer(HBase->RegionServer). Adjust the values of **-Xmx** and **-XX:CMSInitiatingOccupancyFraction** of the GC_OPTS parameter by referring to the Note.

NOTE

1. Suggestions on GC parameter configurations for HMaster
   - Set **-Xms** and **-Xmx** to the same value to prevent JVM from dynamically adjusting the heap memory size and affecting performance.
   - Set **-XX:NewSize** to the value of **-XX:MaxNewSize**, which is one eighth of **-Xmx**.
   - For large-scale HBase clusters with a large number of regions, increase values of **GC_OPTS** parameters for HMaster. Specifically, set **-Xmx** to 4 GB if the number of regions is less than 100,000. If the number of regions is more than 100,000, set -Xmx to be greater than or equal to 6 GB. For each increased 35,000 regions, increase the value of **-Xmx** by 2 GB. The maximum value of **-Xmx** is 32 GB.
2. Suggestions on GC parameter configurations for RegionServer
   - Set **-Xms** and **-Xmx** to the same value to prevent JVM from dynamically adjusting the heap memory size and affecting performance.
   - Set **-XX:NewSize** to one eighth of **-Xmx**.
   - Set the memory for RegionServer to be greater than that for HMaster. If sufficient memory is available, increase the heap memory.
   - Set **-Xmx** based on the machine memory size. Specifically, set **-Xmx** to 32 GB if the machine memory is greater than 200 GB, to 16 GB if the machine memory is greater than 128 GB and less than 200 GB, and to 8 GB if the machine memory is less than 128 GB. When **-Xmx** is set to 32 GB, a RegionServer node supports 2000 regions and 200 hotspot regions.
   - **XX:CMSInitiatingOccupancyFraction** to be less than and equal to **85**, and it is calculated as follows: 100 x (hfile.block.cache.size + hbase.regionserver.global.memstore.size)

Step 5 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager interface of active and standby clusters, choose **O&M** > **Log** > **Download**.

Step 7 In the **Service** drop-down list box, select **HBase** in the required cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected fault logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.153 ALM-19008 Heap Memory Usage of the HBase Process Exceeds the Threshold

Description

The system checks the HBase service status every 30 seconds. The alarm is generated when the heap memory usage of an HBase service exceeds the threshold (90% of the maximum memory).

NOTE

If the multi-instance function is enabled in the cluster and multiple HBase service instances are installed, you need to determine the HBase service instance where the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the HBase1 service is unavailable, **ServiceName=HBase1** is displayed in **Location**, and the operation object in the procedure needs to be changed from HBase to HBase1.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
19008	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.

Impact on the System

If the available HBase heap memory is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The heap memory of the HBase service is overused or the heap memory is inappropriately allocated.

Procedure

Check heap memory usage.

Step 1 On the FusionInsight Manager portal, click **O&M** > **Alarm** > **Alarms** and select the alarm whose **ID** is **19008**. Then check the role name in **Location** and confirm the IP adress of the instance.

- If the role for which the alarm is generated is HMaster, go to **Step 2**.
- If the role for which the alarm is generated is RegionServer, go to **Step 3**.

Step 2 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase** > **Instance** and click the HMaster for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **CPU and Memory** > **HMaster Heap Memory Usage and Direct Memory Usage Statistics** and click **OK**, check whether the used heap memory of the HBase service reaches 90% of the maximum heap memory specified for HBase.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 3 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase** > **Instance** and click the RegionServer for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **CPU and Memory** > **RegionServer Heap Memory Usage and Direct Memory Usage Statistics** and click **OK**, check whether the used heap memory of the HBase service reaches 90% of the maximum heap memory specified for HBase.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase** > **Configurations**, and click **All Configurations**. Choose **HMaster/RegionServer** > **System**. Increase the value of **-Xmx** in **GC_OPTS** by referring to the Note.

NOTE

1. Suggestions on GC parameter configurations for HMaster
   - Set **-Xms** and **-Xmx** to the same value to prevent JVM from dynamically adjusting the heap memory size and affecting performance.
   - Set **-XX:NewSize** to the value of **-XX:MaxNewSize**, which is one eighth of **-Xmx**.
   - For large-scale HBase clusters with a large number of regions, increase values of **GC_OPTS** parameters for HMaster. Specifically, set **-Xmx** to 4 GB if the number of regions is less than 100,000. If the number of regions is more than 100,000, set -Xmx to be greater than or equal to 6 GB. For each increased 35,000 regions, increase the value of **-Xmx** by 2 GB. The maximum value of **-Xmx** is 32 GB.
2. Suggestions on GC parameter configurations for RegionServer
   - Set **-Xms** and **-Xmx** to the same value to prevent JVM from dynamically adjusting the heap memory size and affecting performance.
   - Set **-XX:NewSize** to one eighth of **-Xmx**.
   - Set the memory for RegionServer to be greater than that for HMaster. If sufficient memory is available, increase the heap memory.
   - Set **-Xmx** based on the machine memory size. Specifically, set **-Xmx** to 32 GB if the machine memory is greater than 200 GB, to 16 GB if the machine memory is greater than 128 GB and less than 200 GB, and to 8 GB if the machine memory is less than 128 GB. When **-Xmx** is set to 32 GB, a RegionServer node supports 2000 regions and 200 hotspot regions.

Step 5 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 7 Select **HBase** in the required cluster from the **Service** drop-down list.

Step 8 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected fault logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.154 ALM-19009 Direct Memory Usage of the HBase Process Exceeds the Threshold

Description

The system checks the HBase service status every 30 seconds. The alarm is generated when the direct memory usage of an HBase service exceeds the threshold (90% of the maximum memory).

The alarm is cleared when the direct memory usage is less than the threshold.

NOTE

If the multi-instance function is enabled in the cluster and multiple HBase service instances are installed, you need to determine the HBase service instance where the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the HBase1 service is unavailable, **ServiceName=HBase1** is displayed in **Location**, and the operation object in the procedure needs to be changed from HBase to HBase1.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
19009	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.

Impact on the System

If the available HBase direct memory is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The direct memory of the HBase service is overused or the direct memory is inappropriately allocated.

Procedure

Check direct memory usage.

Step 1 On the FusionInsight Manager portal, click **O&M** > **Alarm** > **Alarms** and select the alarm whose **ID** is **19009**. Check the **RoleName** in **Location** and confirm the IP address of **HostName**.

- If the role for which the alarm is generated is HMaster, go to **Step 2**.
- If the role for which the alarm is generated is RegionServer, go to **Step 3**.

Step 2 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase** > **Instance** and click the HMaster for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **CPU and Memory** > **HMaster Heap Memory Usage and Direct Memory Usage Statistics** and click **OK** to check whether the used direct memory of the HBase service reaches 90% of the maximum direct memory specified for HBase.

- If yes, go to **Step 4**.
- If no, go to **Step 8**.

Step 3 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase** > **Instance** and click the RegionServer for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the **Chart** area and choose **Customize** > **CPU and Memory** > **RegionServer Heap Memory Usage and Direct Memory Usage Statistics** and click **OK** to check whether the used direct memory of the HBase service reaches 90% of the maximum direct memory specified for HBase.

- If yes, go to **Step 4**.
- If no, go to **Step 8**.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase** > **Configurations**, and click **All Configurations**. Choose **HMaster/RegionServer** > **System** and check whether **XX:MaxDirectMemorySize** exists in **GC_OPTS**.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 On the FusionInsight Manager portal, choose **Cluster** > *Nameof the desired cluster* > **Services** > **HBase** > **Configurations**, and click **All Configurations**. Choose **HMaster/RegionServer** > **System** and delete **XX:MaxDirectMemorySize** from **GC_OPTS**.

Step 6 Check whether the **ALM-19008 Heap Memory Usage of the HBase Process Exceeds the Threshold** alarm is generated.

If yes, handle the alarm by referring to **ALM-19008 Heap Memory Usage of the HBase Process Exceeds the Threshold**.

If no, go to **Step 8**.

Step 7 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Collect fault information.

Step 8 On the FusionInsight Manager interface of active and standby clusters, choose **O&M** > **Log** > **Download**.

Step 9 In the **Service** in the required cluster drop-down list box, select **HBase**.

Step 10 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact the O&M personnel and send the collected fault logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.155 ALM-19011 RegionServer Region Number Exceeds the Threshold

Description

The system checks the number of regions on each RegionServer in each HBase service instance every 30 seconds. The region number is displayed on the HBase service monitoring page and RegionServer role monitoring page. This alarm is generated when the number of regions on a RegionServer exceeds the threshold (default value: 2000) for 20 consecutive times. The threshold can be changed by choosing **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **HBase**. This alarm is cleared when the number of regions is less than or equal to the threshold.

NOTE

If the multi-instance function is enabled in the cluster and multiple HBase service instances are installed, you need to determine the HBase service instance where the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the HBase1 service is unavailable, **ServiceName=HBase1** is displayed in **Location**, and the operation object in the procedure needs to be changed from HBase to HBase1.

Attribute

Alarm ID	Alarm Severity	Auto Clear
19011	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The data read/write performance of HBase is affected when the number of regions on a RegionServer exceeds the threshold.

Possible Causes

- The RegionServer region distribution is unbalanced.
- The HBase cluster scale is too small.

Procedure

View alarm location information.

Step 1 On the FusionInsight Manager home page, choose **O&M** > **Alarm** > **Alarms**, select this alarm, and view the service instance and host name in **Location**.

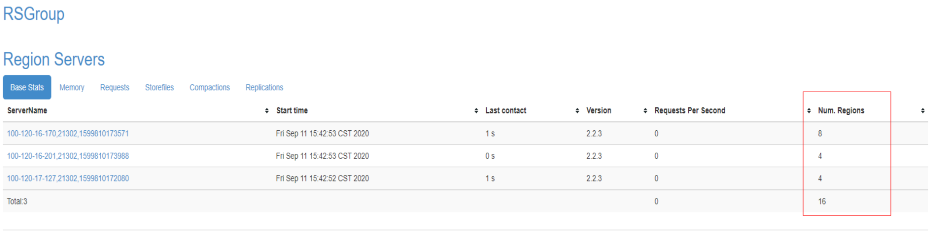
Step 2 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services**, click the HBase service instance for which the alarm is generated, and click **HMaster(Active)**. On the displayed WebUI of the HBase instance, check whether the region distribution on the RegionServer is balanced.

NOTE

By default, the **admin** user does not have the permissions to manage other components. If the page cannot be opened or the displayed content is incomplete when you access the native UI of a component due to insufficient permissions, you can manually create a user with the permissions to manage that component.

- If yes, go to **Step 9**.
- If no, go to **Step 3**.

Figure 9-30 WebUI of HBase instance



RSGroup

Region Servers

ServerName	Start time	Last contact	Version	Requests Per Second	Num. Regions
100-120-16-170,21302,1599810173571	Fri Sep 11 15:42:53 CST 2020	1 s	2.2.3	0	8
100-120-16-201,21302,1599810173988	Fri Sep 11 15:42:53 CST 2020	0 s	2.2.3	0	4
100-120-17-127,21302,1599810172080	Fri Sep 11 15:42:52 CST 2020	1 s	2.2.3	0	4
Total:3				0	16

Enable load balancing.

Step 3 Log in to the node where the HBase client is located as user **root**. Go to the client installation directory, and set environment variables.

cd *client installation directory*

source bigdata_env

If the cluster adopts the security mode, perform security authentication. Specifically, run the **kinit hbase** command and enter the password as prompted (obtain the password from the administrator).

Step 4 Run the following commands to go to the HBase shell command window and check whether the load balancing function is enabled.

hbase shell

balancer_enabled

- If yes, go to **Step 6**.
- If no, go to **Step 5**.

Step 5 On the HBase shell command window, run the following commands to enable the load balancing function and check whether the function is enabled.

balance_switch true

balancer_enabled

Step 6 On the HBase shell command window, run the **balancer** command to manually trigger the load balancing function.

NOTE

You are advised to enable and manually trigger the load balancing function during off-peak hours.

Step 7 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase**, and click **HMaster(Active)**. On the displayed WebUI of the HBase instance, refresh the page and check whether the region distribution is balanced.

- If yes, go to **Step 8**.
- If no, go to **Step 21**.

Step 8 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Delete unwanted HBase tables.

NOTE

Exercise caution when deleting data to ensure data is deleted correctly.

Step 9 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase**, and click **HMaster(Active)**. On the displayed WebUI of the HBase instance, view tables stored in the HBase service instance and record unwanted tables that can be deleted.

Step 10 On the HBase shell command window, run the **disable** command and **drop** command to delete the table to decrease the number of regions.

disable '*name of the table to be deleted***'**

drop '*name of the table to be deleted***'**

Step 11 On the HBase shell command window, run the following command to check whether the load balancing function is enabled.

balancer_enabled

- If yes, go to **Step 13**.
- If no, go to **Step 12**.

Step 12 On the HBase shell command window, run the following commands to enable the load balancing function and confirm that the function is enabled.

balance_switch true

balancer_enabled

Step 13 On the HBase shell command window, run the **balancer** command to manually trigger the load balancing function.

Step 14 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase**, and click **HMaster(Active)**. On the displayed WebUI of the HBase instance, refresh the page and check whether the region distribution is balanced.

- If yes, go to **Step 15**.
- If no, go to **Step 21**.

Step 15 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 16**.

Adjust the threshold.

Step 16 On the FusionInsight Manager home page, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **HBase** > **Regions(RegionServer)**, select the applied rule, and click **Modify** to check whether the threshold is proper.

- If it is excessively small, increase the threshold as required and go to **Step 17**.
- If it is proper, go to **Step 18**.

Step 17 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 18**.

Perform system capacity expansion.

Step 18 Add nodes to the HBase cluster and add RegionServer instances to the nodes. Then enable and manually trigger the load balancing function.

Step 19 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services**, click the HBase service instance for which the alarm is generated, and click **HMaster(Active)**. On the displayed WebUI of the HBase instance, refresh the page and check whether the region distribution is balanced.

- If yes, go to **Step 20**.
- If no, go to **Step 21**.

Step 20 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 21**.

Collect fault information.

Step 21 On the FusionInsight Manager home page of the active and standby clusters, choose **O&M**> **Log** > **Download**.

Step 22 Select **HBase** in the required cluster from the **Service**.

Step 23 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 24 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.156 ALM-19012 HBase System Table Directory or File Lost

Description

The system checks whether HBase directories and files exist on the HDFS every 120 seconds. This alarm is generated when the system detects that the files or directories do not exist. This alarm is cleared when the files or directories are restored.

The HBase directories and files are as follows:

- Directory of the namespace **hbase** on the HDFS
- **hbase.version** file
- Directory of the table **hbase:meta** on the HDFS, .tableinfo file, and .regioninfo file
- Directory of the table **hbase:namespace** on the HDFS, .tableinfo file, and .regioninfo file
- Directory of the table **hbase:hindex** on the HDFS, .tableinfo file, and .regioninfo file
- Directory of the **hbase:acl** table on the HDFS, .tableinfo, and .regioninfo file (This table does not exist in the common mode cluster by default.)

NOTE

If the multi-instance function is enabled in the cluster and multiple HBase service instances are installed, you need to determine the HBase service instance where the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the HBase1 service is unavailable, **ServiceName=HBase1** is displayed in **Location**, and the operation object in the procedure needs to be changed from HBase to HBase1.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
19012	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The HBase service fails to restart or start.

Possible Causes

Files or directories on the HDFS are missing.

Procedure

Locate the alarm cause.

Step 1 On the FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. Click this alarm and check whether **Alarm Cause** indicates unknown errors.

- If yes, go to **Step 4**.
- If no, go to **Step 2**

Step 2 On the FusionInsight Manager home page, choose **O&M** > **Backup and Restoration** > **Backup Management**. Check whether there are success records of the backup task named **default** or other HBase metadata backup tasks that have been successfully executed.

- If yes, go to **Step 3**.
- If no, go to **Step 4**.

Step 3 Use the latest backup metadata to restore the metadata of the HBase service.

Collect fault information.

Step 4 On the FusionInsight Manager page of the active and standby clusters, choose **O&M** > **Log** > **Download**.

Step 5 In the **Service** area, select faulty HBase services in the required cluster.

Step 6 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 7 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.157 ALM-19013 Duration of Regions in transaction State Exceeds the Threshold

Description

The system checks the number of regions in transaction state on HBase every 300 seconds. This alarm is generated when the system detects that the duration of regions in transaction state exceeds the threshold for two consecutive times. This alarm is cleared when all timeout regions are restored.

NOTE

If the multi-instance function is enabled in the cluster and multiple HBase service instances are installed, you need to determine the HBase service instance where the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the HBase1 service is unavailable, **ServiceName=HBase1** is displayed in **Location**, and the operation object in the procedure needs to be changed from HBase to HBase1.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
19013	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

Some data in the table gets lost or becomes unavailable.

Possible Causes

- Compaction is permanently blocked.
- The HDFS files are abnormal.

Procedure

Locate the alarm cause.

Step 1 On the FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**, select this alarm, and view the **HostName** and **RoleName** in **Location**.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services > HBase**, Click the drop-down menu in the chartarea and choose **Customize > Service** >

Region in transaction count to view **Region in transaction count over threshold**. Check whether the monitoring item detects a value in three consecutive detection periods. (The default threshold is 60 seconds.)

- If yes, go to **Step 3**.
- If no, go to **Step 7**.

Step 3 Choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase** > **HMaster (Active)** > **Tables** to check whether the regions of only one table transaction status time out.

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 4 Run the **hbase hbck** command on the client and check whether the error message "No table descriptor file under hdfs://hacluster/hbase/data/default/table" is displayed.

- If yes, go to **Step 5**.
- If no, go to **Step 7**.

Step 5 Log in to the client as user **root**. Run the following command:

cd *client installation directory*

source bigdata_env

If the cluster is in security mode, run the **kinit hbase** command

Log in to the HMaster WebUI, choose **Procedure & Locks** in the navigation tree, and check whether any process ID is in the **Waiting** state in **Procedures**. If yes, run the following command to release the procedure lock:

hbase hbck -j *client installation directory***/HBase/hbase/tools/hbase-hbck2-*.jar bypass -o** *pid*

Check whether the state is in the **Bypass** state. If the procedure on the UI is always in **RUNNABLE(Bypass)** state, perform an active/standby switchover. Run the **assigns** command to bring the region online again.

hbase hbck -j *client installation directory***/HBase/hbase/tools/hbase-hbck2-*.jar assigns -o** *regionName*

Step 6 Repeat **Step 4**. Run the **hbase hbck** command on the client and check whether the error message "No table descriptor file under hdfs://hacluster/hbase/data/default/table" is displayed.

- If yes, go to **Step 7**.
- If no, no further action is required.

Collect fault information.

Step 7 On the FusionInsight Manager page of the active and standby clusters, choose **O&M** > **Log** > **Download**.

Step 8 In the **Service** area, select faulty HBase services in the required cluster.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.158 ALM-19014 Capacity Quota Usage on ZooKeeper Exceeds the Threshold Severely

Description

The system checks the ZNode usage of the HBase service every 120 seconds. This alarm is generated when the ZNode capacity usage of the HBase service exceeds the critical alarm threshold (90% by default).

This alarm is cleared when the ZNode capacity usage is less than the critical alarm threshold.

NOTE

If the multi-instance function has been enabled in the cluster and multiple HBase services have been installed, determine the HBase service for which the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the value of **ServiceName** is **HBase-1**, change the operation object in the procedure from **HBase** to **HBase-1**.

Attribute

Alarm ID	Alarm Severity	Auto Clear
19014	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Threshold	Specifies the threshold for which the alarm is generated.

Impact on the System

This alarm indicates that the capacity usage of the ZNode of HBase has exceeded the threshold severely. As a result, the write request of the HBase service fails.

Possible Causes

- DR is configured for HBase, and data synchronization fails or is slow in DR.
- A large number of WAL files are being split in the HBase cluster.

Procedure

Check the capacity configuration and usage of ZNodes.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**, select the alarm whose ID is **19014**, and view the threshold in **Additional Information**.

Step 2 Log in to the HBase client as user **root**. Run the following command to go to the client installation directory:

cd *Client installation directory*

Run the following command to set environment variables:

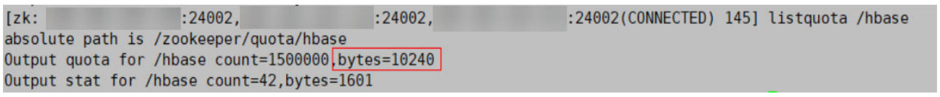
source bigdata_env

If the cluster uses the security mode, run the following command to perform security authentication:

kinit hbase

Enter the password as prompted (obtain the password from the MRS cluster administrator).

Step 3 Run the **hbase zkcli** command to log in to the ZooKeeper client and run the **listquota /hbase** command to check the ZNode capacity quota of the HBase service. The ZNode root directory in the command is specified by the **zookeeper.znode.parent** parameter of the HBase service. The marked area in the following figure shows the capacity configuration of the root ZNode of the HBase service.



```
[zk: [redacted]:24002,[redacted]:24002,[redacted]:24002(CONNECTED) 145] listquota /hbase
absolute path is /zookeeper/quota/hbase
Output quota for /hbase count=1500000,bytes=10240
Output stat for /hbase count=42,bytes=1601
```

Step 4 Run the **getusage /hbase/splitWAL** command to check the capacity usage of the ZNode. Check whether the ratio of **Data size** to the ZNode capacity quota is close to the alarm threshold.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. Check whether the alarm whose ID is **12007**, **19000**, or **19013** and the **ServiceName** in **Location** is the current HBase service exists.

- If yes, click **View Help** next to the alarm and rectify the fault by referring to the help document. Then, go to **Step 8**.
- If no, go to **Step 9**.

Step 6 Run the **getusage /hbase/replication** command to check the capacity usage of the ZNode. Check whether the ratio of **Data size** to the ZNode capacity quota is close to the alarm threshold.

- If yes, go to **Step 7**.
- If no, go to **Step 9**.

Step 7 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. Check whether the alarm whose ID is **19006** and **ServiceName** in **Location** is the current HBase service exists.

- If yes, click **View Help** next to the alarm and rectify the fault by referring to the help document. Then, go to **Step 8**.
- If no, go to **Step 9**.

Step 8 Check whether the alarm is cleared five minutes later.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect the fault information.

Step 9 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 10 Expand the drop-down list next to the **Service** field. In the **Services** dialog box that is displayed, select **HBase** for the target cluster.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.159 ALM-19015 Quantity Quota Usage on ZooKeeper Exceeds the Threshold

Description

The system checks the ZNode usage of the HBase service every 120 seconds. This alarm is generated when the system detects that the ZNode quantity usage of the HBase service exceeds the alarm threshold (75% by default).

This alarm is cleared when the ZNode quantity usage is less than the alarm threshold.

NOTE

If the multi-instance function has been enabled in the cluster and multiple HBase services have been installed, determine the HBase service for which the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the value of **ServiceName** is **HBase-1**, change the operation object in the procedure from **HBase** to **HBase-1**.

Attribute

Alarm ID	Alarm Severity	Auto Clear
19015	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Threshold	Specifies the threshold for which the alarm is generated.

Impact on the System

This alarm indicates that the ZNode quantity usage in the HBase service has exceeded the threshold. If this alarm is not handled in a timely manner, the problem severity may be escalated to **Critical**, affecting data writing.

Possible Causes

- DR is configured for HBase, and data synchronization fails or is slow in DR.
- A large number of WAL files are being split in the HBase cluster.

Procedure

Check the quantity quota and usage of ZNodes.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**, select the alarm whose ID is **19015**, and view the threshold in **Additional Information**.

Step 2 Log in to the HBase client as user **root**. Run the following command to go to the client installation directory:

cd *Client installation directory*

Run the following command to set environment variables:

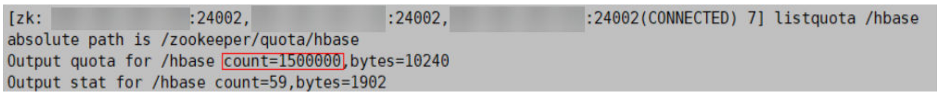
source bigdata_env

If the cluster uses the security mode, run the following command to perform security authentication:

kinit hbase

Enter the password as prompted (obtain the password from the MRS cluster administrator).

Step 3 Run the **hbase zkcli** command to log in to the ZooKeeper client and run the **listquota /hbase** command to check the ZNode quantity quota of the HBase service. The ZNode root directory in the command is specified by the **zookeeper.znode.parent** parameter of the HBase service. The marked area in the following figure shows the quantity quota configuration of the root ZNode of the HBase service.



```
[zk: [redacted]:24002,[redacted]:24002,[redacted]:24002(CONNECTED) 7] listquota /hbase
absolute path is /zookeeper/quota/hbase
Output quota for /hbase count=1500000,bytes=10240
Output stat for /hbase count=59,bytes=1902
```

Step 4 Run the **getusage /hbase/splitWAL** command to check the ZNode quantity usage and check whether the ratio of **Node count** in the command output to the ZNode quantity quota is close to the alarm threshold.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. Check whether the alarm whose ID is **12007**, **19000**, or **19013** and the **ServiceName** in **Location** is the current HBase service exists.

- If yes, click **View Help** next to the alarm and rectify the fault by referring to the help document. Then, go to **Step 8**.
- If no, go to **Step 9**.

Step 6 Run the **getusage /hbase/replication** command to check the ZNode quantity usage and check whether the ratio of **Node count** in the command output to the ZNode quantity quota is close to the alarm threshold.

- If yes, go to **Step 7**.
- If no, go to **Step 9**.

Step 7 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. Check whether the alarm whose ID is **19006** and **ServiceName** in **Location** is the current HBase service exists.

- If yes, click **View Help** next to the alarm and rectify the fault by referring to the help document. Then, go to **Step 8**.
- If no, go to **Step 9**.

Step 8 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect the fault information.

Step 9 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 10 Expand the **Service** drop-down list, and select **HBase** for the target cluster.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.160 ALM-19016 Quantity Quota Usage on ZooKeeper Exceeds the Threshold Severely

Description

The system checks the ZNode usage of the HBase service every 120 seconds. This alarm is generated when the znode usage of the HBase service exceeds the critical alarm threshold (90% by default).

This alarm is cleared when the quantity usage of the ZNode is less than the critical alarm threshold.

NOTE

If the multi-instance function has been enabled in the cluster and multiple HBase services have been installed, determine the HBase service for which the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the value of **ServiceName** is **HBase-1**, change the operation object in the procedure from **HBase** to **HBase-1**.

Attribute

Alarm ID	Alarm Severity	Auto Clear
19016	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.

Name	Meaning
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Threshold	Specifies the threshold for which the alarm is generated.

Impact on the System

This alarm indicates that the quantity usage of the ZNode of HBase has exceeded the threshold severely. As a result, the write request of the HBase service fails.

Possible Causes

- DR is configured for HBase, and data synchronization fails or is slow in DR.
- A large number of WAL files are being split in the HBase cluster.

Procedure

Check the quantity quota and usage of ZNodes.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**, select the alarm whose ID is **19016**, and view the threshold in **Additional Information**.

Step 2 Log in to the HBase client as user **root**. Run the following command to go to the client installation directory:

cd *Client installation directory*

Run the following command to set environment variables:

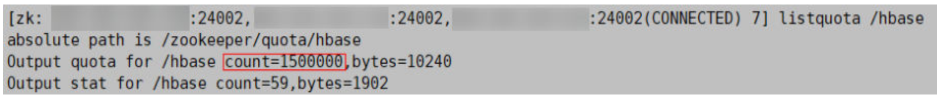
source bigdata_env

If the cluster uses the security mode, run the following command to perform security authentication:

kinit hbase

Enter the password as prompted (obtain the password from the MRS cluster administrator).

Step 3 Run the **hbase zkcli** command to log in to the ZooKeeper client and run the **listquota /hbase** command to check the ZNode quantity quota of the HBase service. The ZNode root directory in the command is specified by the **zookeeper.znode.parent** parameter of the HBase service. The marked area in the following figure shows the quantity configuration of the root ZNode of the HBase service.



```
[zk: [redacted]:24002,[redacted]:24002,[redacted]:24002(CONNECTED) 7] listquota /hbase
absolute path is /zookeeper/quota/hbase
Output quota for /hbase count=1500000,bytes=10240
Output stat for /hbase count=59,bytes=1902
```

Step 4 Run the **getusage /hbase/splitWAL** command to check the ZNode usage and check whether the ratio of **Node count** in the command output to the znode quantity quota is close to the alarm threshold.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. Check whether the alarm whose ID is **12007**, **19000**, or **19013** and the **ServiceName** in **Location** is the current HBase service exists.

- If yes, click **View Help** next to the alarm and rectify the fault by referring to the help document. Then, go to **Step 8**.
- If no, go to **Step 9**.

Step 6 Run the **getusage /hbase/replication** command to check the ZNode usage and check whether the ratio of **Node count** in the command output to the ZNode quantity quota is close to the alarm threshold.

- If yes, go to **Step 7**.
- If no, go to **Step 9**.

Step 7 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. Check whether the alarm whose ID is **19006** and **ServiceName** in **Location** is the current HBase service exists.

- If yes, click **View Help** next to the alarm and rectify the fault by referring to the help document. Then, go to **Step 8**.
- If no, go to **Step 9**.

Step 8 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect the fault information.

Step 9 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 10 Expand the **Service** drop-down list, and select **HBase** for the target cluster.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.161 ALM-19017 Capacity Quota Usage on ZooKeeper Exceeds the Threshold

Description

The system checks the ZNode usage of the HBase service every 120 seconds. This alarm is generated when the system detects that the ZNodes capacity usage of the HBase service exceeds the alarm threshold (75% by default).

This alarm is cleared when the capacity usage of the ZNode capacity is less than the threshold.

NOTE

If the multi-instance function has been enabled in the cluster and multiple HBase services have been installed, determine the HBase service for which the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the value of **ServiceName** is **HBase-1**, change the operation object in the procedure from **HBase** to **HBase-1**.

Attribute

Alarm ID	Alarm Severity	Auto Clear
19017	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Threshold	Specifies the threshold for which the alarm is generated.

Impact on the System

This alarm indicates that the ZNodes capacity usage in the HBase service has exceeded the threshold. If this alarm is not handled in a timely manner, the problem severity may be escalated to **Critical**, affecting data writing.

Possible Causes

- DR is configured for HBase, and data synchronization fails or is slow in DR.
- A large number of WAL files are being split in the HBase cluster.

Procedure

Check the capacity configuration and usage of ZNodes.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**, select the alarm whose ID is **19017**, and view the threshold in **Additional Information**.

Step 2 Log in to the HBase client as user **root**. Run the following command to go to the client installation directory:

cd *Client installation directory*

Run the following command to set environment variables:

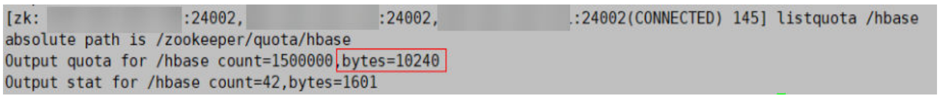
source bigdata_env

If the cluster uses the security mode, run the following command to perform security authentication:

kinit hbase

Enter the password as prompted (obtain the password from the MRS cluster administrator).

Step 3 Run the **hbase zkcli** command to log in to the ZooKeeper client and run the **listquota /hbase** command to check the ZNode quantity quota of the HBase service. The ZNode root directory in the command is specified by the **zookeeper.znode.parent** parameter of the HBase service. The marked area in the following figure shows the quantity configuration of the root ZNode of the HBase service.



```
[zk: [redacted]:24002,[redacted]:24002,[redacted].:24002(CONNECTED) 145] listquota /hbase
absolute path is /zookeeper/quota/hbase
Output quota for /hbase count=1500000,bytes=10240
Output stat for /hbase count=42,bytes=1601
```

Step 4 Run the **getusage /hbase/splitWAL** command to check the capacity usage of the ZNode. Check whether the ratio of **Data size** to the ZNode capacity quota is close to the alarm threshold.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 On FusionInsight Manager, check whether the alarm whose ID is **12007**, **19000**, or **19013** and **ServiceName** in **Location** is the current HBase service exists.

- If yes, click **View Help** next to the alarm and rectify the fault by referring to the help document. Then, go to **Step 8**.
- If no, go to **Step 7**.

Step 6 Run the **getusage /hbase/replication** command to check the capacity usage of the ZNode. Check whether the ratio of **Data size** to the ZNode capacity quota is close to the alarm threshold.

- If yes, go to **Step 7**.

- If no, go to **Step 9**.

Step 7 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. Check whether the alarm whose ID is **19006** and **ServiceName** in **Location** is the current HBase service exists.

- If yes, click **View Help** next to the alarm and rectify the fault by referring to the help document. Then, go to **Step 8**.
- If no, go to **Step 9**.

Step 8 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect the fault information.

Step 9 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 10 Expand the **Service** drop-down list, and select **HBase** for the target cluster.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.162 ALM-19018 HBase Compaction Queue Size Exceeds the Threshold

Description

The system checks the HBase compaction queue size every 300 seconds. This alarm is generated when the compaction queue size exceeds the alarm threshold (**100** by default). This alarm is cleared when the compaction queue size is less than the threshold.

NOTE

If the multi-instance function has been enabled in the cluster and multiple HBase services have been installed, determine the HBase service for which the alarm is generated based on the value of **ServiceName** in **Location**. For example, if the value of **ServiceName** is **HBase-1**, change the operation object in the procedure from **HBase** to **HBase-1**.

Attribute

Alarm ID	Alarm Severity	Auto Clear
19018	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The cluster performance may deteriorate, affecting data read and write.

Possible Causes

- The number of HBase RegionServers is too small.
- There are excessive regions on a single RegionServer of HBase.
- The HBase RegionServer heap size is small.
- Resources are insufficient.
- Related parameters are not configured properly.

Procedure

Check whether related parameters are properly configured.

Step 1 Log in to FusionInsight Manager and choose **O&M** > **Alarm** > **Alarms**. On the page that is displayed, check whether the alarm whose **Alarm ID** is **19008** or **19011** exists.

- If yes, click **View Help** next to the alarm and rectify the fault by referring to the help document. Then, go to **Step 3**.
- If no, go to **Step 2**.

Step 2 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase**. On the page that is disaplyed, click the **Configuration** tab then the **All Configurations** sub-tab, search for **hbase.hstore.compaction.min**, **hbase.hstore.compaction.max**, **hbase.regionserver.thread.compaction.small**, and **hbase.regionserver.thread.compaction.throttle**, and set them to larger values.

Step 3 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Collect the fault information.

Step 4 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 5 Expand the **Service** drop-down list, and select **HBase** for the target cluster.

Step 6 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 7 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.163 ALM-19019 Number of HBase HFiles to Be Synchronized Exceeds the Threshold

Description

The system checks the number of HFiles to be synchronized by the RegionServer of each HBase service instance every 30 seconds. This indicator can be viewed on the RegionServer role monitoring page. This alarm is generated when the number of HFiles to be synchronized on a RegionServer exceeds the threshold (exceeding 128 for 20 consecutive times by default). To change the threshold, choose **O&M** > **Alarm** > **Threshold Configuration** > *Name of the desired cluster* > **HBase** . This alarm is cleared when the number of HFiles to be synchronized is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
19019	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

If the number of HFiles to be synchronized by a RegionServer exceeds the threshold, the number of ZNodes used by HBase exceeds the threshold, affecting the HBase service status.

Possible Causes

- The network is abnormal.
- The RegionServer region distribution is unbalanced.
- The HBase service scale of the standby cluster is too small.

Procedure

View alarm location information.

Step 1 Log in to FusionInsight Manager and choose **O&M**. In the navigation pane on the left, choose **Alarm** > **Alarms**. On the page that is displayed, locate the row containing the alarm whose **Alarm ID** is **19019**, and view the service instance and host name in **Location**.

Check the network connection between RegionServers on active and standby clusters.

Step 2 Run the **ping** command to check whether the network connection between the faulty RegionServer node and the host where RegionServer of the standby cluster resides is normal.

- If yes, go to **Step 5**.
- If no, go to **Step 3**.

Step 3 Contact the network administrator to restore the network.

Step 4 After the network recovers, check whether the alarm is cleared.

- If yes, no further action is required.

- If no, go to **Step 5**.

Check the RegionServer region distribution in the active cluster.

Step 5 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase**. Click **HMaster(Active)** to go to the web UI of the HBase instance and check whether regions are evenly distributed on the Region Server.

Step 6 Log in to the faulty RegionServer node as user **omm**.

Step 7 Run the following commands to go to the client installation directory and set the environment variable:

cd *Client installation directory*

source bigdata_env

If the cluster uses the security mode, perform security authentication. Run the **kinit hbase** command and enter the password as prompted (obtain the password from the MRS cluster administrator).

Step 8 Run the following commands to check whether the load balancing function is enabled.

hbase shell

balancer_enabled

- If yes, go to **Step 10**.
- If no, go to **Step 9**.

Step 9 Run the following commands in HBase Shell to enable the load balancing function and check whether the function is enabled.

balance_switch true

balancer_enabled

Step 10 Run the **balancer** command to manually trigger the load balancing function.

NOTE

You are advised to enable and manually trigger the load balancing function during off-peak hours.

Step 11 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 12**.

Check the HBase service scale of the standby cluster.

Step 12 Expand the HBase cluster, add a node, and add a RegionServer instance on the node. Then, perform **Step 6** to **Step 10** to enable the load balancing function and manually trigger it.

Step 13 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase**. Click **HMaster(Active)** to go to the web UI of the HBase instance, refresh the page, and check whether regions are evenly distributed.

- If yes, go to **Step 14**.

- If no, go to **Step 15**.

Step 14 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 15**.

Collect the fault information.

Step 15 On FusionInsight Manager of the standby cluster, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 16 Expand the **Service** drop-down list, and select **HBase** for the target cluster.

Step 17 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 18 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.164 ALM-19020 Number of HBase WAL Files to Be Synchronized Exceeds the Threshold

Description

The system checks the number of WAL files to be synchronized by the RegionServer of each HBase service instance every 30 seconds. This indicator can be viewed on the RegionServer role monitoring page. This alarm is generated when the number of WAL files to be synchronized on a RegionServer exceeds the threshold (exceeding 128 for 20 consecutive times by default). To change the threshold, choose **O&M** > **Alarm** > **Threshold Configuration** > *Name of the desired cluster* > **HBase** . This alarm is cleared when the number of WAL files to be synchronized is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
19020	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

If the number of WAL files to be synchronized by a RegionServer exceeds the threshold, the number of ZNodes used by HBase exceeds the threshold, affecting the HBase service status.

Possible Causes

- The network is abnormal.
- The RegionServer region distribution is unbalanced.
- The HBase service scale of the standby cluster is too small.

Procedure

View alarm location information.

Step 1 Log in to FusionInsight Manager and choose **O&M**. In the navigation pane on the left, choose **Alarm** > **Alarms**. On the page that is displayed, locate the row containing the alarm whose **Alarm ID** is **19020**, and view the service instance and host name in **Location**.

Check the network connection between RegionServers on active and standby clusters.

Step 2 Run the **ping** command to check whether the network connection between the faulty RegionServer node and the host where RegionServer of the standby cluster resides is normal.

- If yes, go to **Step 5**.
- If no, go to **Step 3**.

Step 3 Contact the network administrator to restore the network.

Step 4 After the network recovers, check whether the alarm is cleared.

- If yes, no further action is required.

- If no, go to **Step 5**.

Check the RegionServer region distribution in the active cluster.

Step 5 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase**. Click **HMaster(Active)** to go to the web UI of the HBase instance and check whether regions are evenly distributed on the Region Server.

Step 6 Log in to the faulty RegionServer node as user **omm**.

Step 7 Run the following commands to go to the client installation directory and set the environment variable:

cd *Client installation directory*

source bigdata_env

If the cluster uses the security mode, perform security authentication. Run the **kinit hbase** command and enter the password as prompted (obtain the password from the MRS cluster administrator).

Step 8 Run the following commands to check whether the load balancing function is enabled.

hbase shell

balancer_enabled

- If yes, go to **Step 10**.
- If no, go to **Step 9**.

Step 9 Run the following commands in HBase Shell to enable the load balancing function and check whether the function is enabled.

balance_switch true

balancer_enabled

Step 10 Run the **balancer** command to manually trigger the load balancing function.

NOTE

You are advised to enable and manually trigger the load balancing function during off-peak hours.

Step 11 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 12**.

Check the HBase service scale of the standby cluster.

Step 12 Expand the HBase cluster, add a node, and add a RegionServer instance on the node. Then, perform **Step 6** to **Step 10** to enable the load balancing function and manually trigger it.

Step 13 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **HBase**. Click **HMaster(Active)** to go to the web UI of the HBase instance, refresh the page, and check whether regions are evenly distributed.

- If yes, go to **Step 14**.

- If no, go to **Step 15**.

Step 14 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 15**.

Collect the fault information.

Step 15 On FusionInsight Manager of the standby cluster, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 16 Expand the **Service** drop-down list, and select **HBase** for the target cluster.

Step 17 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 18 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.165 ALM-20002 Hue Service Unavailable

Description

This alarm is generated when the Hue service is unavailable. The system checks the Hue service status every 60 seconds.

This alarm is cleared when the Hue service is normal.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
20002	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The system cannot provide data loading, query, and extraction services.

Possible Causes

- The internal KrbServer service on which the Hue service depends is abnormal.
- The internal DBService service on which the Hue service depends is abnormal.
- The network connection to the DBService is abnormal.

Procedure

Check whether the KrbServer is abnormal.

Step 1 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services**. In the service list, check whether the **KrbServer** running status is **Normal**.

- If yes, go to **Step 4**.
- If no, go to **Step 2**.

Step 2 Restart the KrbServer service.

Step 3 Wait several minutes, and check whether **Hue Service Unavailable** is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check whether the DBService is abnormal.

Step 4 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 5 In the service list, check whether the **DBService** running status is **Normal**.

- If yes, go to **Step 8**.
- If no, go to **Step 6**.

Step 6 Restart the DBService.

NOTE

To restart the service, enter the FusionInsight Manager administrator password.

Step 7 Wait several minutes, and check whether **Hue Service Unavailable** is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Check whether the network connection to the DBService is normal.

Step 8 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Hue** > **Instance**, record the IP address of the active Hue.

Step 9 Log in to the active Hue.

Step 10 Run the **ping** command to check whether communication between the host that runs the active Hue and the hosts that run the DBService is normal. (Obtain the IP addresses of the hosts that run the DBService in the same way as that for obtaining the IP address of the active Hue.)

- If yes, go to **Step 13**.
- If no, go to **Step 11**.

Step 11 Contact the administrator to restore the network.

Step 12 Wait several minutes, and check whether **Hue Service Unavailable** is cleared.

- If yes, no further action is required.
- If no, go to **Step 13**.

Collect fault information.

Step 13 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 14 Select the following nodes in the required cluster from the **Service** drop-down list:

- Hue
- Controller

Step 15 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 16 On the FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Hue**.

Step 17 Choose **More** > **Restart Service**, and click **OK**.

Step 18 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 19**.

Step 19 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.166 ALM-24000 Flume Service Unavailable

Description

The alarm module checks the Flume service status every 180 seconds. This alarm is generated if the Flume service is abnormal.

This alarm is automatically cleared after the Flume service recovers.

Attribute

Alarm ID	Alarm Severity	Auto Clear
24000	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

Flume cannot work and data transmission is interrupted.

Possible Causes

All Flume instances are faulty.

Procedure

Step 1 Log in to a Flume node as user **omm** and run the **ps -ef|grep "flume.role=server"** command to check whether the Flume process exists on the node.

- If yes, go to **Step 3**.

- If no, restart the faulty Flume node or Flume service and go to **Step 2**.

Step 2 In the alarm list, check whether alarm "Flume Service Unavailable" is cleared.

- If yes, no further action is required.
- If no, go to **Step 3**.

Collect the fault information.

Step 3 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 4 Expand the **Service** drop-down list, and select **Flume** for the target cluster.

Step 5 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 6 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.167 ALM-24001 Flume Agent Exception

Description

The Flume agent instance for which the alarm is generated cannot be started. This alarm is generated when the Flume agent process is faulty (The system checks in every 5 seconds.) or Flume agent fails to start (The system reporting alarms immediately).

This alarm is cleared when the Flume agent process recovers, Flume agent starts successfully and the alarm handling is completed.

Attribute

Alarm ID	Alarm Severity	Auto Clear
24001	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
AgentId	Specifies the ID of the agent for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The Flume agent instance for which the alarm is generated cannot provide services properly, and the data transmission tasks of the instance are temporarily interrupted. Real-time data is lost during real-time data transmission.

Possible Causes

- The JAVA_HOME directory does not exist or the Java permission is incorrect.
- The Flume agent directory permission is incorrect.
- Flume agent fails to start.

Procedure

Check whether the JAVA_HOME directory exists or whether the JAVA permission is correct.

Step 1 Log in to the host for which the alarm is generated as user **root**.

Step 2 Run the following command to obtain the installation directory of the Flume client for which the alarm is generated: (The value of **AgentId** can be obtained from **Location** of the alarm.)

ps -ef|grep *AgentId* **| grep -v grep | awk -F 'conf-file ' '{print $2}' | awk -F 'fusioninsight' '{print $1}'**

Step 3 Run the **su -** *Flume installation user* command to switch to the Flume installation user and run the **cd** *Flume client installation directory***/fusioninsight-flume-1.9.0/conf/** command to go to the Flume configuration directory.

Step 4 Run the **cat ENV_VARS | grep JAVA_HOME** command.

Step 5 Check whether the **JAVA_HOME** directory exists. If the command output in **Step 4** is not empty and **ll $JAVA_HOME/** is not empty, the **JAVA_HOME** directory exists.

- If yes, go to **Step 7**.

- If no, go to **Step 6**.

Step 6 Specify a correct **JAVA_HOME** directory.

Step 7 Run the **$JAVA_HOME/bin/java -version** command to check whether the Flume agent running user has the Java execution permission. If the Java version is displayed in the command output, the Java permission meets the requirement. Otherwise, the Java permission does not meet the requirement.

- If yes, go to **Step 9**.
- If no, go to **Step 8**.

NOTE

JAVA_HOME is the environment variable exported during Flume client installation. You can also go to *Flume client installation directory***/fusioninsight-flume-1.9.0/conf** and run the **cat ENV_VARS | grep JAVA_HOME** command to view the variable value.

Step 8 Run the **chmod 750 $JAVA_HOME/bin/java** command to grant the Java execution permission to the Flume agent running user.

Check the directory permission of the Flume agent.

Step 9 Log in to the host for which the alarm is generated as user **root**.

Step 10 Run the following command to switch to the Flume agent installation directory:

cd *Flume client installation directory***/fusioninsight-flume-1.9.0/conf/**

Step 11 Run the **ls -al * -R** command to check whether any file owner is the user who running the Flume agent.

- If yes, go to **Step 12**.
- If no, run the **chown** command to change the file owner to the user who runs the Flume agent.

Check the Flume agent configuration.

Step 12 Run the **cat properties.properties | grep spooldir** and **cat properties.properties | grep TAILDIR** commands to check whether the Flume source type is spoolDir or tailDir. If any command output is displayed, the Flume source type is spoolDir or tailDir.

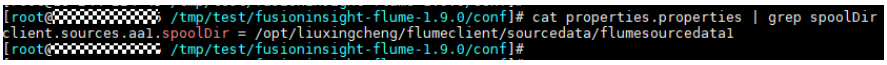
- If yes, go to **Step 13**.
- If no, go to **Step 17**.

Step 13 Check whether the data monitoring directory exists.

- If yes, go to **Step 15**.
- If no, go to **Step 14**.

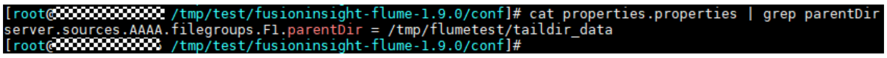
NOTE

Run the **cat properties.properties | grep spoolDir** command to view the spoolDir monitoring directory.



```
[root@[redacted] /tmp/test/fusioninsight-flume-1.9.0/conf]# cat properties.properties | grep spoolDir
client.sources.aal.spoolDir = /opt/liuxingcheng/flumeclient/sourcedata/flumesourcedatal
[root@[redacted] /tmp/test/fusioninsight-flume-1.9.0/conf]#
```

Run the **cat properties.properties | grep parentDir** command to view the tailDir monitoring directory.



```
[root@[redacted] /tmp/test/fusioninsight-flume-1.9.0/conf]# cat properties.properties | grep parentDir
server.sources.AAAA.filegroups.F1.parentDir = /tmp/flumetest/taildir_data
[root@[redacted] /tmp/test/fusioninsight-flume-1.9.0/conf]#
```

Step 14 Specify a correct data monitoring directory.

Step 15 Check whether the Flume agent user has the read, write, and execute permissions on the monitoring directory specified in **Step 13**.

- If yes, go to **Step 17**.
- If no, go to **Step 16**.

NOTE

Go to the monitoring directory as the Flume running user. If files can be created, the Flume running user has the read, write, and execute permissions on the monitoring directory.

Step 16 Run the **chmod 777** *Flume monitoring directory* command to grant the Flume agent running user the read, write, and execute permissions on the monitoring directory specified in **Step 13**.

Step 17 Check whether the components connected to the Flume sink are in safe mode.

- If yes, go to **Step 18**.
- If no, go to **Step 23**.

NOTE

If the sinks in the **properties.properties** configuration file are the HDFS sink and HBase sink, and the configuration file contains a keytab file, the components connected to the Flume sink are in safe mode.

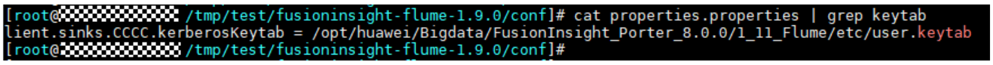
If the sink in the **properties.properties** configuration file is the kafka sink and ***.security.protocol** is set to **SASL_PLAINTEXT** or **SASL_SSL**, Kafka connected to the Flume sink is in safe mode.

Step 18 Run the **ll** *ketab path* command to check whether the keytab authentication path specified by the ***.kerberosKeytab** parameter in the configuration file exists.

- If yes, go to **Step 20**.
- If no, go to **Step 19**.

NOTE

To view the ketab path, run the **cat properties.properties | grep keytab** command.



```
[root@[redacted] /tmp/test/fusioninsight-flume-1.9.0/conf]# cat properties.properties | grep keytab
lient.sinks.CCCC.kerberosKeytab = /opt/huawei/Bigdata/FusionInsight_Porter_8.0.0/1_11_Flume/etc/user.keytab
[root@[redacted] /tmp/test/fusioninsight-flume-1.9.0/conf]#
```

Step 19 Change the value of **kerberosKeytab** in **Step 18** to the custom keytab path and go to **Step 21**.

Step 20 Perform **Step 18** to check whether the Flume agent running user has the permission to access the keytab authentication file. If the keytab path is returned, the user has the permission. Otherwise, the user does not have the permission.

- If yes, go to **Step 22**.
- If no, go to **Step 21**.

Step 21 Run the **chmod 755** *ketab file* command to grant the read permission on the keytab file specified in **Step 19**, and restart the Flume process.

Step 22 Check whether the alarm is cleared.

- If yes, no further action is required.

- If no, go to **Step 23**.

Collect the fault information.

Step 23 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 24 Expand the **Service** drop-down list, and select **Flume** for the target cluster.

Step 25 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 26 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.168 ALM-24003 Flume Client Connection Interrupted

Description

The alarm module monitors the port connection status on the Flume server. This alarm is generated if the Flume server fails to receive a connection message from the Flume client in three consecutive minutes.

This alarm is cleared after the Flume server receives a connection message from the Flume client.

Attribute

Alarm ID	Alarm Severity	Auto Clear
24003	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
Client IP Address	Specifies the IP address of the Flume client.

Name	Meaning
Client Name	Specifies the agent name of the Flume client.
Sink Name	Specifies the sink name of Flume Agent.

Impact on the System

The communication between the Flume client and the server fails. The Flume client cannot send data to the Flume server.

Possible Causes

- The network connection between the Flume client and the server is faulty.
- The Flume client's process is abnormal.
- The Flume client is incorrectly configured.

Procedure

Check the network connection between the Flume client and the server.

Step 1 Log in to the host whose IP address is specified by **Flume ClientIP** in the alarm information as user **root**.

Step 2 Run the **ping** *Flume server IP address* command to check whether the network connection between the Flume client and the server is normal.

- If yes, go to **Step 3**.
- If no, go to **Step 11**.

Check whether the Flume client's process is normal.

Step 3 Log in to the host whose IP address is specified by **Flume ClientIP** in the alarm information as user **root**.

Step 4 Run the **ps -ef|grep flume |grep client** command to check whether the Flume client process exists.

- If yes, go to **Step 5**.
- If no, go to **Step 11**.

Check the Flume client configuration.

Step 5 Log in to the host whose IP address is specified by **Flume ClientIP** in the alarm information as user **root**.

Step 6 Run the **cd** *Flume client installation directory*/**fusioninsight-flume-1.9.0/conf/** command to go to Flume's configuration directory.

Step 7 Run the **cat properties.properties** command to query the current configuration file of the Flume client.

Step 8 Check whether the **properties.properties** file is correctly configured according to the configuration description of the Flume agent.

- If yes, go to **Step 9**.
- If no, go to **Step 11**.

Step 9 Modify the **properties.properties** configuration file.

Check whether the alarm is cleared.

Step 10 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 11**.

Collect the fault information.

Step 11 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 12 Expand the **Service** drop-down list, and select **Flume** for the target cluster.

Step 13 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 14 Collect logs in the **/var/log/Bigdata/flume-client** directory on the Flume client using a transmission tool.

Step 15 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.169 ALM-24004 Exception Occurs When Flume Reads Data

Description

The alarm module monitors the status of Flume Source. This alarm is generated immediately when the duration in which Source fails to read the data exceeds the threshold.

The default threshold is **0**, indicating that the threshold is disabled. You can change the threshold by modifying the **properties.properties** file in the **conf** directory. Specifically, modify the **NoDatatime** parameter of required the source.

The alarm is cleared when Source reads the data and the alarm handling is complete.

Attribute

Alarm ID	Alarm Severity	Auto Clear
24004	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
AgentId	Specifies the ID of the agent for which the alarm is generated.
ComponentType	Specifies the component type for which the alarm is generated.
ComponentName	Specifies the component name for which the alarm is generated.

Impact on the System

If data is found in the data source and Flume Source continuously fails to read data, the data collection is stopped.

Possible Causes

- Flume Source is faulty, so data cannot be sent.
- The network is faulty, so the data cannot be sent.

Procedure

Check whether Flume Source is faulty.

Step 1 Open the **properties.properties** configuration file on the local PC, search for **keyword type = spooldir** in the file, and check whether the Flume source type is spoolDir.

- If yes, go to **Step 2**.
- If no, go to **Step 3**.

Step 2 View the spoolDir directory to check whether all files are already transferred.

- If yes, no further action is required.

- If no, go to **Step 5**.

NOTE

The monitoring directory of spooDir is specified by the **.spoolDir** parameter in the **properties.properties** configuration file. If all files in the monitoring directory have been transferred, the file name extension of all files in the monitoring directory is **.COMPLETED**.

Step 3 Open the **properties.properties** configuration file on the local PC, search for **org.apache.flume.source.kafka.KafkaSource** in the file, and check whether the Flume source type is Kafka.

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Step 4 Check whether the topic data configured by Kafka Source has been used up.

- If yes, no further action is required.
- If no, go to **Step 5**.

Step 5 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Flume** > **Instance**.

Step 6 Go to the Flume instance page of the faulty node to check whether the indicator **Source Speed Metrics** in the alarm is 0.

- If yes, go to **Step 11**.
- If no, go to **Step 7**.

Check the network connection between the faulty node and the node that corresponds to the Flume Source IP address.

Step 7 Open the **properties.properties** configuration file on the local PC, search for **type = avro** in the file, and check whether the Flume source type is Avro.

- If yes, go to **Step 8**.
- If no, go to **Step 11**.

Step 8 Log in to the faulty node as user **root**, and run the **ping** *IP address of the Flume source* command to check whether the peer host can be pinged successfully.

- If yes, go to **Step 11**.
- If no, go to **Step 9**.

Step 9 Contact the network administrator to restore the network.

Step 10 In the alarm list, check whether the alarm is cleared after a period.

- If yes, no further action is required.
- If no, go to **Step 11**.

Collect the fault information.

Step 11 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 12 Expand the **Service** drop-down list, and select **Flume** for the target cluster.

Step 13 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 14 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.170 ALM-24005 Exception Occurs When Flume Transmits Data

Description

The alarm module monitors the capacity status of Flume Channel. The alarm is generated immediately when the duration that Channel is fully occupied exceeds the threshold or the number of times that Source fails to send data to Channel exceeds the threshold.

The default threshold is **10**. You can change the threshold by modifying the **channelfullcount** parameter of the related channel in the **properties.properties** configuration file in the **conf** directory.

The alarm is cleared when the space of Flume Channel is released and the alarm handling is complete.

Attribute

Alarm ID	Alarm Severity	Auto Clear
24005	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Name	Meaning
AgentId	Specifies the ID of the agent for which the alarm is generated.
ComponentType	Specifies the type of the component for which the alarm is generated.
ComponentName	Specifies the component for which the alarm is generated.

Impact on the System

If the disk usage of Flume Channel increases continuously, the time required for importing data to a specified destination prolongs. When the disk usage of Flume Channel reaches 100%, the Flume agent process pauses.

Possible Causes

- Flume Sink is faulty, so the data cannot be sent.
- The network is faulty, so the data cannot be sent.

Procedure

Check whether Flume Sink is faulty.

Step 1 Open the **properties.properties** configuration file on the local PC, search for **type = hdfs** in the file, and check whether the Flume sink type is HDFS.

- If yes, go to **Step 2**.
- If no, go to **Step 3**.

Step 2 On FusionInsight Manager, check whether **HDFS Service Unavailable** alarm is generated in the alarm list and whether the HDFS service is stopped in the service list.

- If the alarm is reported, clear it according to the handling suggestions of ALM-14000 HDFS Service Unavailable; if the HDFS service is stopped, start it. Then, go to **Step 7**.
- If no, go to **Step 7**.

Step 3 Open the **properties.properties** configuration file on the local PC, search for **type = hbase** in the file, and check whether the Flume sink type is HBase.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 On FusionInsight Manager, check whether **HBase Service Unavailable** alarm is generated in the alarm list and whether the HBase service is stopped in the service list.

- If the alarm is reported, clear it according to the handling suggestions of ALM-19000 HBase Service Unavailable; if the HBase service is stopped, start it. Then, go to **Step 7**.

- If no, go to **Step 7**.

Step 5 Open the **properties.properties** configuration file on the local PC, search for **org.apache.flume.sink.kafka.KafkaSink** in the file, and check whether the Flume sink type is Kafka.

- If yes, go to **Step 6**.
- If no, go to **Step 9**.

Step 6 On FusionInsight Manager, check whether **Kafka Service Unavailable** alarm is generated in the alarm list and whether the Kafka service is stopped in the service list.

- If the alarm is reported, clear it according to the handling suggestions of ALM-38000 Kafka Service Unavailable; if the Kafka service is stopped, start it. Then, go to **Step 7**.
- If no, go to **Step 7**.

Step 7 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Flume** > **Instance**.

Step 8 Go to the Flume instance page of the faulty node to check whether the indicator **Sink Speed Metrics** is 0.

- If yes, go to **Step 13**.
- If no, go to **Step 9**.

Check the network connection between the faulty node and the node that corresponds to the Flume Sink IP address.

Step 9 Open the **properties.properties** configuration file on the local PC, search for **type = avro** in the file, and check whether the Flume sink type is Avro.

- If yes, go to **Step 10**.
- If no, go to **Step 13**.

Step 10 Log in to the faulty node as user **root**, and run the **ping** *IP address of the Flume sink* command to check whether the peer host can be pinged successfully.

- If yes, go to **Step 13**.
- If no, go to **Step 11**.

Step 11 Contact the network administrator to restore the network.

Step 12 In the alarm list, check whether the alarm is cleared after a period.

- If yes, no further action is required.
- If no, go to **Step 13**.

Collect the fault information.

Step 13 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 14 Expand the **Service** drop-down list, and select **Flume** for the target cluster.

Step 15 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 16 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.171 ALM-24006 Heap Memory Usage of Flume Server Exceeds the Threshold

Description

The system checks the heap memory usage of the Flume service every 60 seconds. This alarm is generated when the heap memory usage of the Flume instance exceeds the threshold (95% of the maximum memory) for 10 consecutive times. This alarm is cleared when the heap memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
24006	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Heap memory overflow may cause service breakdown.

Possible Causes

The heap memory of the Flume instance is overused or the heap memory is inappropriately allocated.

Procedure

Check the heap memory usage.

Step 1 Log in to FusionInsight Manager and choose **O&M**. In the navigation pane on the left, choose **Alarm** > **Alarms**. On the page that is displayed, locate the row containing **Flume Heap Memory Usage Exceeds the Threshold**, and view the **Location** information. Check the name of the host for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > *Name of the target cluster* > **Services** > **Flume**. On the page that is displayed, click the **Instance** tab. On the displayed tab page, select the role corresponding to the host name for which the alarm is generated and select **Customize** from the drop-down list in the upper right corner of the chart area. Choose **Agent** and select **Flume Heap Memory Resource Percentage**. Then, click **OK**.

Step 3 Check whether the heap memory used by Flume reaches the threshold (95% of the maximum heap memory by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Service** > **Flume** > **Configuration**. On the page that is displayed, click **All Configurations** and choose **Flume** > **System**. Set **-Xmx** in the **GC_OPTS** parameter to a larger value based on site requirements and save the configuration.

NOTE

If this alarm is generated, the heap memory configured for the Flume server is insufficient for data transmission. You are advised to change the heap memory to: Channel capacity x Maximum size of a single data record x Number of channels. Note that the value of **xmx** cannot exceed the remaining memory of the node.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Flume** for the target cluster.

Step 8 Click in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.172 ALM-24007 Flume Server Direct Memory Usage Exceeds the Threshold

Description

The system checks the direct memory usage of the Flume service every 60 seconds. This alarm is generated when the direct memory usage of the Flume instance exceeds the threshold (80% of the maximum memory) for five consecutive times. This alarm is cleared when the Flume direct memory usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
24007	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Direct memory overflow may cause service breakdown.

Possible Causes

The direct memory of the Flume process is overused or the direct memory is inappropriately allocated.

Procedure

Check the direct memory usage.

Step 1 Log in to FusionInsight Manager and choose **O&M**. In the navigation pane on the left, choose **Alarm** > **Alarms**. On the page that is displayed, locate the row containing **Flume Direct Memory Usage Exceeds the Threshold**, and view the **Location** information. Check the name of the host for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > *Name of the target cluster* > **Services** > **Flume**. On the page that is displayed, click the **Instance** tab. On the displayed tab page, select the role corresponding to the host name for which the alarm is generated and select **Customize** from the drop-down list in the upper right corner of the chart area. Choose **Agent** and select **Flume Direct Memory Resource Percentage**. Then, click **OK**.

Step 3 Check whether the direct memory used by Flume reaches the threshold (80% of the maximum direct memory by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Service** > **Flume** > **Configuration**. On the page that is displayed, click **All Configurations** and choose **Flume** > **System**. Set **-XX:MaxDirectMemorySize** in the **GC_OPTS** parameter to a larger value based on site requirements and save the configuration.

NOTE

If this alarm is generated, the direct memory size configured for the Flume server instance cannot meet service requirements. You are advised to change the value of **-XX:MaxDirectMemorySize** to twice the current direct memory size or change the value based on site requirements.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Flume** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.173 ALM-24008 Flume Server Non Heap Memory Usage Exceeds the Threshold

Description

The system checks the non-heap memory usage of the Flume service every 60 seconds. This alarm is generated when the non-heap memory usage of the Flume instance exceeds the threshold (80% of the maximum memory) for five consecutive times. This alarm is cleared when the non-heap memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
24008	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Non-heap memory overflow may cause service breakdown.

Possible Causes

The non-heap memory of the Flume instance is overused or the non-heap memory is inappropriately allocated.

Procedure

Check non-heap memory usage.

Step 1 Log in to FusionInsight Manager and choose **O&M**. In the navigation pane on the left, choose **Alarm** > **Alarms**. On the page that is displayed, locate the row containing **Flume Non Heap Memory Usage Exceeds the Threshold**, and view the **Location** information. Check the name of the host for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > *Name of the target cluster* > **Services** > **Flume**. On the page that is displayed, click the **Instance** tab. On the displayed tab page, select the role corresponding to the host name for which the alarm is generated and select **Customize** from the drop-down list in the upper right corner of the chart area. Choose **Agent** and select **Flume Non Heap Memory Resource Percentage**. Then, click **OK**.

Step 3 Check whether the non-heap memory used by Flume reaches the threshold (80% of the maximum non-heap memory by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Service** > **Flume** > **Configuration**. On the page that is displayed, click **All Configurations** and choose **Flume** > **System**. Set **-XX: MaxPermSize** in the **GC_OPTS** parameter to a larger value based on site requirements and save the configuration.

NOTE

If this alarm is generated, the non-heap memory size configured for the Flume server instance cannot meet service requirements. You are advised to change the value of **-XX:MaxPermSize** to twice the current non-heap memory size or change the value based on site requirements.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Flume** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.174 ALM-24009 Flume Server Garbage Collection (GC) Time Exceeds the Threshold

Description

The system checks the GC duration of the Flume process every 60 seconds. This alarm is generated when the GC duration of the Flume process exceeds the threshold (12 seconds by default) for five consecutive times. This alarm is cleared when the GC duration is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
24009	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Flume data transmission efficiency decreases.

Possible Causes

The heap memory of the Flume process is overused or inappropriately allocated, causing frequent occurrence of the GC process.

Procedure

Check the GC duration.

Step 1 Log in to FusionInsight Manager and choose **O&M**. In the navigation pane on the left, choose **Alarm** > **Alarms**. On the page that is displayed, locate the row containing **GC Duration Exceeds the Threshold**, and view the **Location** information. Check the name of the host for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > *Name of the target cluster* > **Services** > **Flume**. On the page that is displayed, click the **Instance** tab. On the displayed tab page, select the role corresponding to the host name for which the alarm is generated and select **Customize** from the drop-down list in the upper right corner of the chart area. Choose **Agent** and select **Garbage Collection (GC) Duration of Flume**. Then, click **OK**.

Step 3 Check whether the GC duration of the Flume process collected every minute exceeds the threshold (12 seconds by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Service** > **Flume** > **Configuration**. On the page that is displayed, click **All Configurations** and choose **Flume** > **System**. Set **-Xmx** in the **GC_OPTS** parameter to a larger value based on site requirements and save the configuration.

NOTE

If this alarm is generated, the heap memory configured for the Flume server is insufficient for data transmission. You are advised to change the heap memory to: Channel capacity x Maximum size of a single data record x Number of channels. Note that the value of **xmx** cannot exceed the remaining memory of the node.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Flume** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.175 ALM-25000 LdapServer Service Unavailable

Description

The system checks the LdapServer service status every 30 seconds. This alarm is generated when the system detects that both the active and standby LdapServer services are abnormal.

This alarm is cleared when the system detects that one or two LdapServer services are normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
25000	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

When this alarm is generated, no operation can be performed for the KrbServer users and LdapServer users in the cluster. For example, users, user groups, or roles cannot be added, deleted, or modified, and user passwords cannot be changed on the FusionInsight Manager portal. The authentication for existing users in the cluster is not affected.

Possible Causes

- The node where the LdapServer service locates is faulty.
- The LdapServer process is abnormal.

Procedure

Check whether the nodes where the two SlapdServer instances of the LdapServer service are located are faulty.

Step 1 On FusionInsight Manager, choose **Cluster >** *Name of the desired cluster* **> Services** > **LdapServer** > **Instance** to go to the LdapServer instance page to obtain the host name of the node where the two SlapdServer instances locates.

Step 2 Choose **O&M > Alarm > Alarms**. On the **Alarm** page of the FusionInsight Manager system, check whether any alarm of **Node Fault** exists.

- If yes, go to **Step 3**.
- If no, go to **Step 6**.

Step 3 Check whether the host name in the alarm is consistent with the **Step 1** host name.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 Handle the alarm according to "ALM-12006 Node Fault".

Step 5 Check whether **LdapServer Service Unavailable** is cleared in the alarm list.

- If yes, no further action is required.
- If no, go to **Step 10**.

Check whether the LdapServer process is normal.

Step 6 Choose **O&M > Alarm > Alarms**. On the **Alarm** page of the FusionInsight Manager system, check whether any alarm of **Process Fault** exists.

- If yes, go to **Step 7**.
- If no, go to **Step 10**.

Step 7 Check whether the service and host name in the alarm are consistent with the LdapServer service and host name.

- If yes, go to **Step 8**.
- If no, go to **Step 10**.

Step 8 Handle the alarm according to "ALM-12007 Process Fault".

Step 9 Check whether **LdapServer Service Unavailable** is cleared in the alarm list.

- If yes, no further action is required.
- If no, go to **Step 10**.

Collect fault information.

Step 10 On the FusionInsight Manager, choose **O&M** > **Log > Download**.

Step 11 Select **LdapServer** in the required cluster from the **Service**.

Step 12 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 13 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.176 ALM-25004 Abnormal LdapServer Data Synchronization

Description

The system checks the LdapServer data every 30 seconds. This alarm is generated when the data on the active and standby LdapServers of Manager is inconsistent for 12 consecutive times. This alarm is cleared when the data on the active and standby LdapServers is consistent.

The system checks the LdapServer data every 30 seconds. This alarm is generated when the LdapServer data in the cluster is inconsistent with that on Manager for 12 consecutive times. This alarm is cleared when the data is consistent.

Attribute

Alarm ID	Alarm Severity	Auto Clear
25004	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

LdapServer data inconsistency occurs because the LdapServer data in Manager is damaged or the LdapServer data in the cluster is damaged. The LdapServer process with damaged data cannot provide services externally, and the authentication functions of Manager and the cluster are affected.

Possible Causes

- The network of the node where the LdapServer process locates is faulty.
- The LdapServer process is abnormal.
- The OS restart damages data on LdapServer.

Procedure

Check whether the network where the LdapServer nodes reside is faulty.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms**. Record the IP address of HostName in the alarm locating information as IP1 (if multiple alarms exist, record the IP addresses as IP1, IP2, and IP3 respectively).

Step 2 Contact O&M personnel and log in to the nodes corresponding to IP 1. Run the ping command to check whether the IP address of the management plane of the active OMS node can be pinged.

- If yes, go to **Step 4**.
- If no, go to **Step 3**.

Step 3 Contact the network administrator to recover the network and check whether **Abnormal LdapServer Data Synchronization** is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Check whether the LdapServer processes are normal.

Step 4 On the **Alarm** page of FusionInsight Manager, check whether the **OLdap Resource Abnormal** exists.

- If yes, go to **Step 5**.
- If no, go to **Step 7**.

Step 5 Clear the alarm by following the steps provided in "ALM-12004 OLdap Resource Abnormal".

Step 6 Check whether **Abnormal LdapServer Data Synchronization** is cleared in the alarm list.

- If yes, no further action is required.
- If no, go to **Step 7**.

Step 7 On the **Alarm** page of FusionInsight Manager, check whether **Process Faul**t is generated for the LdapServer service.

- If yes, go to **Step 8**.
- If no, go to **Step 10**.

Step 8 Handle the alarm according to "ALM-12007 Process Fault".

Step 9 Check whether **Abnormal LdapServer Data Synchronization** is cleared.

- If yes, no further action is required.
- If no, go to **Step 10**.

Check whether the LdapServer processes are normal.

Step 10 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. Record the IP address of HostName in the alarm locating information as "IP1" (if multiple alarms exist, record the IP addresses as "IP1", "IP2", and "IP3" respectively). Choose **Cluster** > *Name of the desired cluster* > **Services** > **LdapServer** > **Configurations**. Record the port number of LdapServer as "PORT". (If the IP address in the alarm locating information is the IP address of the standby management node, choose **System** > **OMS** > **oldap** > **Modify Configuration** and record the listening port number of LdapServer.)

Step 11 Log in to the nodes corresponding to IP1 as user **omm**.

Step 12 Run the following command to check whether errors are displayed in the queried information.

ldapsearch -H ldaps://*IP1*:*PORT* **-LLL -x -D cn=root,dc=hadoop,dc=com -W -b ou=Peoples,dc=hadoop,dc=com**

After running the command, enter the **LDAP** administrator password. Contact the system administrator to obtain the password.

- If yes, go to **Step 13**.
- If no, go to **Step 15**.

Step 13 Recover the LdapServer and OMS nodes using data backed up before the alarm is generated.

> NOTE
>
> Use the OMS data and LdapServer data backed up at the same point in time to recover the data. Otherwise, the service and operation may fail. To recover data when services run properly, you are advised to manually back up the latest management data and then recover the data. Otherwise, Manager data produced between the backup point in time and the recovery point in time will be lost.

Step 14 Check whether alarm **Abnormal LdapServer Data Synchronization** is cleared.

- If yes, no further action is required.
- If no, go to **Step 15**.

Collect fault information.

Step 15 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 16 Select **LdapServer** in the required cluster and **OmsLdapServer** from the **Service**.

Step 17 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 18 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.177 ALM-25005 nscd Service Exception

Description

The system checks the status of the nscd service every 60 seconds. This alarm is generated when the nscd process fails to be queried for four consecutive times (three minutes) or users in LdapServer cannot be obtained.

This alarm is cleared when the process is restored and users in LdapServer can be obtained.

Attribute

Alarm ID	Alarm Severity	Auto Clear
25005	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
HostName	Host for which the alarm is generated.

Impact on the System

The alarmed node may not be able to synchronize data from LdapServer. The **id** command may fail to obtain the LDAP data, affecting upper-layer services.

Possible Causes

- The nscd service is not started.
- The network is faulty, and cannot access the LDAP server.
- NameService is abnormal.
- Users cannot be queried because the OS executes commands too slowly.

Procedure

Check whether the nscd service is started.

Step 1 Log in to FusionInsight Manager and choose **O&M** > **Alarm** > **Alarms**. Record the IP address of **HostName** in **Location** of the alarm as **IP1** (if multiple alarms exist, record the IP addresses as **IP1**, **IP2**, and **IP3** respectively).

Step 2 Contact the O&M personnel to access the node using IP1 as user **root**. Run the **ps -ef | grep nscd** command on the node and check whether the **/usr/sbin/nscd** process is started.

- If yes, go to **Step 5**.
- If no, go to **Step 3**.

Step 3 Run the **service nscd restart** command as user **root** to restart the nscd service. Then run the **ps -ef | grep nscd** command to check whether the nscd service is started.

- If yes, go to **Step 4**.
- If no, go to **Step 15**.

Step 4 Wait for 5 minutes and run the ps -ef | grep nscd command again as user root. Check whether the service exists.

- If yes, go to **Step 11**.
- If no, go to **Step 15**.

Check whether the network is faulty, and whether the LDAP server can be accessed.

Step 5 Log in to the alarmed node as user **root** and run the **ping** command to check whether the network connectivity between this node and the LdapServer node. is normal.

- If yes, go to **Step 6**.
- If no, contact network administrators to troubleshoot the fault.

Check whether the NameService is normal.

Step 6 Log in to the alarmed node as user **root**. Run the **cat /etc/nsswitch.conf** command to check whether the **passwd**, **group**, **services**, **netgroup**, and **aliases** of NameService are correctly configured.

The correct parameter configurations are as follows:

passwd: **compat ldap**; **group**: **compat ldap**; **services**: **files ldap**; **netgroup**: **files ldap**; **aliases**: **files ldap**

- If yes, go to **Step 7**.
- If no, go to **Step 9**.

Step 7 Log in to the alarmed node as user **root**. Run the **cat /etc/nscd.conf** command to check whether the **enable-cache passwd**, **positive-time-to-live passwd**, **enable-cache group**, and **positive-time-to-live group** in the configuration file are correctly configured.

The correct parameter configurations are as follows:

enable-cache passwd: **yes**; **positive-time-to-live passwd**: **600**; **enable-cache group**: **yes**; **positive-time-to-live group**: **3600**

- If yes, go to **Step 8**.
- If no, go to **Step 10**.

Step 8 Run the **/usr/sbin/nscd -i group** and **/usr/sbin/nscd -i passwd** commands as user **root**. Wait for 2 minutes and run the **id admin** and **id backup/manager** commands to check whether results can be queried.

- If yes, go to **Step 11**.
- If no, go to **Step 15**.

Step 9 Run the **vi /etc/nsswitch.conf** command as user **root**. Correct the configurations in **Step 6** and save the file. Run the **service nscd restart** command to restart the nscd service. Wait for 2 minutes and run the **id admin** and **id backup/manager** commands to check whether results can be queried.

- If yes, go to **Step 11**.
- If no, go to **Step 15**.

Step 10 Run the **vi /etc/nscd.conf** command as user **root**. Correct the configurations in **Step 7** and save the file. Run the **service nscd restart** command to restart the nscd service. Wait for 2 minutes and run the **id admin** and **id backup/manager** commands to check whether results can be queried.

- If yes, go to **Step 11**.
- If no, go to **Step 15**.

Step 11 Log in to the FusionInsight Manager portal. Wait for 5 minutes and check whether the **nscd Service Exception** alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 12**.

Check whether frame freezing occurs when running a command in the operating system.

Step 12 Log in to the faulty node as user **root**, run the **id admin** command, and check whether the command execution takes a long time. If the command execution takes more than 3 seconds, the command execution is deemed to be slow.

- If yes, go to **Step 13**.
- If no, go to **Step 15**.

Step 13 Run the **cat /var/log/messages** command to check whether the nscd frequently restarts or the error information "Can't contact LDAP server" exists.

nscd exception example:

```
Feb 11 11:44:42 10-120-205-33 nscd: nss_ldap: failed to bind to LDAP server ldaps://10.120.205.55:21780: 
Can't contact LDAP server
Feb 11 11:44:43 10-120-205-33 ntpq: nss_ldap: failed to bind to LDAP server ldaps://10.120.205.55:21780: 
Can't contact LDAP server
Feb 11 11:44:44 10-120-205-33 ntpq: nss_ldap: failed to bind to LDAP server ldaps://10.120.205.92:21780: 
Can't contact LDAP server
```

- If yes, go to **Step 14**.
- If no, go to **Step 15**.

Step 14 Run the **vi$BIGDATA_HOME/tmp/random_ldap_ip_order** command to modify the number at the end. If the original number is an odd number, change it to an even number. If the number is an even number, change it to an odd number.

Run the **vi /etc/ldap.conf** command to enter the editing mode, press **Insert** to start editing, and then change the first two IP addresses of the URI configuration item.

After the modification is complete, press **Esc** to exit the editing mode and enter **:wq!** to save the settings and exit.

Run the **service nscd restart** command to restart the nscd service. Wait 5 minutes and run the **id admin** command again. Check whether the command execution is slow.

- If yes, go to **Step 15**.
- If no, log in to other faulty nodes and repeat **Step 12** to **Step 14** to check whether the first LdapServer node in the URI before modifying **/etc/ldap.conf** is faulty. For example, check whether the service IP address is unreachable, the network delay is too long, or other abnormal software is deployed.

Collect the fault information.

Step 15 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 16 Expand the drop-down list next to the **Service** field. In the **Services** dialog box that is displayed, select **LdapClient** for the target cluster.

Step 17 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 18 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.178 ALM-25006 Sssd Service Exception

Description

The system checks the status of the sssd service every 60 seconds. This alarm is generated when the sssd process fails to be queried for four consecutive times (three minutes) or users in LdapServer cannot be obtained.

This alarm is cleared when the process is restored and users in LdapServer can be obtained.

Attribute

Alarm ID	Alarm Severity	Auto Clear
25006	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.

Impact on the System

The alarmed node may not be able to synchronize data from LdapServer. The id command may fail to obtain the LDAP data, affecting upper-layer services.

Possible Causes

- The sssd service is not started or is incorrectly started.
- The network is faulty and cannot access the LDAP server.
- NameService is abnormal.
- Users cannot be queried because the OS executes commands too slowly.

Procedure

Check whether the sssd service is correctly started.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms**. Find the IP address of **HostName** in **Location** of the alarm and record it as IP1 (if multiple alarms exist, record the IP addresses as IP1, IP2, and IP3 respectively).

Step 2 Contact the O&M personnel to access the node using IP1 as user root. Run the **ps -ef | grep sssd** command and check whether the **/usr/sbin/sssd** process is started.

- If the process is started, go to **Step 3**.
- If the process is not started, go to **Step 4**.

Step 3 Check whether the sssd process queried in **Step 2** has three subprocesses.

- If yes, go to **Step 5**.
- If no, go to **Step 4**.

Step 4 Run the **service sssd restart** command as user **roo**t to restart the sssd service. Then run the **ps -ef | grep sssd** command to check whether the sssd process is normal.

In the normal state, the **/usr/sbin/sssd** process has three subprocesses: **/usr/libexec/sssd/sssd_be**, **/usr/libexec/sssd/sssd_nss**, and **/usr/libexec/sssd/sssd_pam**.

- If it exists, go to **Step 9**.
- If it does not exist, go to **Step 13**.

Check whether the LDAP server can be accessed.

Step 5 Log in to the alarmed node as user **root**. Run the **ping** command to check the network connectivity between this node and the LdapServer node.

- If the network is normal, go to **Step 6**.
- If the network is faulty, contact network administrators to troubleshoot the fault.

Check whether NameService is normal.

Step 6 Log in to the alarmed node as user **root**. Run the **cat /etc/nsswitch.conf** command and check the **passwd** and **group** configurations of NameService.

The correct parameter configurations are as follows: **passwd: compat ldap** and **group: compat ldap**.

- If the configurations are correct, go to **Step 7**.
- If the configurations are incorrect, go to **Step 8**.

Step 7 Run the **/usr/sbin/sss_cache -G** and **/usr/sbin/sss_cache -U** commands as user **root**. Wait for 2 minutes and run the **id admin** and **id backup/manager** commands to check whether results can be queried.

- If results are queried, go to **Step 9**.
- If no result is queried, go to **Step 13**.

Step 8 Run the **vi /etc/nsswitch.conf** command as user **root**. Correct the configurations in **Step 6** and save the file. Run the **service sssd restart** command to restart the sssd service. Wait for 2 minutes and run the **id admin** and **id backup/manager** commands to check whether results can be queried.

- If results are queried, go to **Step 9**.
- If no result is queried, go to **Step 13**.

Step 9 Log in to the FusionInsight Manager portal. Wait for 5 minutes and check whether the **sssd Service Exception** alarm is cleared.

- If the alarm is cleared, no further action is required.
- If the alarm persists, go to **Step 10**.

Check whether frame freezing occurs when running a command in the operating system.

Step 10 Log in to the faulty node as user **root**, run the **id admin** command, and check whether the command execution takes a long time. If the command execution takes more than 3 seconds, the command execution is deemed to be slow.

- If yes, go to **Step 11**.
- If no, go to **Step 13**.

Step 11 Run the **cat /var/log/messages** command to check whether the sssd frequently restarts or the error information **Can't contact LDAP server** exists.

sssd restart example:

```
Feb  7 11:38:16 10-132-190-105 sssd[pam]: Shutting down
Feb  7 11:38:16 10-132-190-105 sssd[nss]: Shutting down
Feb  7 11:38:16 10-132-190-105 sssd[nss]: Shutting down
Feb  7 11:38:16 10-132-190-105 sssd[be[default]]: Shutting down
Feb  7 11:38:16 10-132-190-105 sssd: Starting up
Feb  7 11:38:16 10-132-190-105 sssd[be[default]]: Starting up
Feb  7 11:38:16 10-132-190-105 sssd[nss]: Starting up
Feb  7 11:38:16 10-132-190-105 sssd[pam]: Starting up
```

- If yes, go to **Step 12**.
- If no, go to **Step 13**.

Step 12 Run the **vi $BIGDATA_HOME/tmp/random_ldap_ip_order** command to modify the number at the end. If the original number is an odd number, change it to an even number. If the number is an even number, change it to an odd number.

Run the **vi /etc/sssd/sssd.conf** command to reverse the first two IP addresses of the **ldap_uri** configuration item, save the settings, and exit.

Run the **ps -ef | grep sssd** command to query the ID of the sssd process, kill it, and run the **/usr/sbin/sssd -D -f** command to restart the sssd service. Wait 5 minutes and run the **id admin** command again.

Check whether the command execution is slow.

- If yes, go to **Step 13**.
- If no, log in to other faulty nodes and run **Step 10** to **Step 12**. Collect logs and check whether the first ldapserver node in the ldap_uri before modifying **/etc/sssd/sssd.conf** is faulty. For example, check whether the service IP address is unreachable, the network latency is too long, or other abnormal software is deployed.

Collect fault information.

Step 13 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 14 Select **LdapClient** in the required cluster from the **Service**.

Step 15 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 16 Contact the O&M personnel and send the collected fault logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.179 ALM-25500 KrbServer Service Unavailable

Description

The system checks the KrbServer service status every 30 seconds. This alarm is generated when the system detects that the KrbServer service is abnormal.

This alarm is cleared when the system detects that the KrbServer service is normal.

Attribute

Alarm ID	Alarm Severity	Auto Clear
25500	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

When this alarm is generated, no operation can be performed for the KrbServer component in the cluster. The authentication of KrbServer in other components will be affected. The running status of components that depend on KrbServer in the cluster is Bad.

Possible Causes

- The node where the KrbServer service locates is faulty.
- The OLdap service is abnormal.

Procedure

Check whether the node where the KrbServer service locates is faulty.

Step 1 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **KrbServer** > **Instance** to go to the KrbServer instance page to obtain the host name of the node where the KrbServer service locates.

Step 2 On the **Alarm** page of the FusionInsight Manager system, check whether any alarm of **Node Fault** exists.

- If yes, go to **Step 3**.
- If no, go to **Step 6**.

Step 3 Check whether the host name in the alarm is consistent with the **Step 1** host name.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 Handle the alarm according to "ALM-12006 Node Fault".

Step 5 Check whether **KrbServer Service Unavailable** is cleared in the alarm list.

- If yes, no further action is required.
- If no, go to **Step 6**.

Check whether the OLdap service is normal.

Step 6 On the **Alarm** page of the FusionInsight Manager system, check whether any alarm of **OLdap Resource Abnormal** exists.

- If yes, go to **Step 7**.
- If no, go to **Step 9**.

Step 7 Handle the alarm according to "ALM-12004 OLdap Resource Abnormal".

Step 8 Check whether **KrbServer Service Unavailable** is cleared in the alarm list.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect fault information.

Step 9 On the FusionInsight Manager, choose **O&M** > **Log > Download**.

Step 10 Select **KrbServer** in the required cluster from the **Service**.

Step 11 Click in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.180 ALM-26051 Storm Service Unavailable

Description

The system checks the Storm service status every 30 seconds. This alarm is generated when all Nimbus nodes in the cluster are abnormal and the Storm service is unavailable.

This alarm is cleared when the Storm service recovers.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
26051	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The cluster cannot provide the Storm service, and users cannot perform new Storm tasks.

Possible Causes

- The Kerberos cluster is faulty.
- The ZooKeeper cluster is faulty or suspended.
- The active and standby Nimbus nodes in the Storm cluster are abnormal

Procedure

Check the status of the Kerberos cluster. (Skip this step if the normal mode is used.)

Step 1 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services**.

Step 2 Check whether the running status of the Kerberos service is **Normal**.

- If yes, go to **Step 5**.
- If no, go to **Step 3**.

Step 3 See the related maintenance information of **ALM-25500 KrbServer Service Unavailable**.

Step 4 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check the status of the ZooKeeper cluster.

Step 5 Check whether the running status of the ZooKeeper service is **Normal**.

- If yes, go to **Step 8**.
- If no, go to **Step 6**.

Step 6 If ZooKeeper service is stopped, start it, else see the related maintenance information of **ALM-13000 ZooKeeper Service Unavailable**.

Step 7 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Check the status of the active and standby Nimbus nodes.

Step 8 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Storm** > **Nimbus** to go to the Nimbus Instances page.

Step 9 Check whether only one Nimbus node that is in the **Active** state in **Roles**.

- If yes, go to **Step 13**.
- If no, go to **Step 10**.

Step 10 Select two Nimbus role instances, choose **More** > **Restart Instance**, and check whether the instances restart successfully.

- If yes, go to **Step 11**.
- If no, go to **Step 13**.

Step 11 Log in to the FusionInsight Manager portal again, choose **Cluster** > *Name of the desired cluster* > **Services** > **Storm** > **Nimbus** to check whether the running status is **Normal**.

- If yes, go to **Step 12**.
- If no, go to **Step 13**.

Step 12 Wait for 30 seconds and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 13**.

Collecting Fault Information

Step 13 On the FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 14 Select the following nodes in the required cluster from the **Service** drop-down list:

- KrbServer

  NOTE

  KrbServer logs do not need to be downloaded in normal mode.
- ZooKeeper
- Storm

Step 15 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 16 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.181 ALM-26052 Number of Available Supervisors of the Storm Service Is Less Than the Threshold

Description

The system periodically checks the number of available Supervisors every 60 seconds and compares the number of available Supervisors with the threshold. This alarm is generated when the number of available Supervisors is less than the threshold.

You can change the threshold in **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster*.

This alarm is cleared when the number of available Supervisors is greater than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
26052	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

Existing tasks in the cluster cannot be performed. The cluster can receive new Storm tasks, but cannot perform these tasks.

Possible Causes

The status of some Supervisors in the cluster is abnormal.

Procedure

Check the Supervisor status.

Step 1 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Storm** > **Supervisor** to go to the Storm service management page.

Step 2 In **Roles**, check whether any instance whose status is **Faulty** or **Restoring** exists.

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 Select Supervisor role instances whose status is **Faulty** or **Restoring**, choose **More** > **Restart Instance**, and check whether the instances restart successfully.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 Wait for 30 seconds, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

NOTE

Services are interrupted when the Supervisor is being restarted. Then, services are restored after the restarting.

Collect fault information.

Step 5 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 6 Select **Storm** and **ZooKeeper** in the required cluster from the **Service** drop-down list box.

Step 7 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 8 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.182 ALM-26053 Storm Slot Usage Exceeds the Threshold

Description

The system checks the slot usage every 60 seconds and compares the actual slot usage with the threshold. This alarm is generated when the slot usage is greater than the threshold.

You can change the threshold in **O&M** > **Alarm** > **Thresholds**.

This alarm is cleared when the slot usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
26053	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

New Storm tasks cannot be performed.

Possible Causes

- The status of some Supervisors in the cluster is abnormal.
- The status of all Supervisors is normal, but the processing capability is insufficient.

Procedure

Check the Supervisor status.

Step 1 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Storm** > **Instance** to go to the Storm instance management page.

Step 2 Check whether any instance whose status is **Faulty** or **Restoring** exists.

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 Select Supervisor role instances whose status is **Faulty** or **Restoring**, choose **More** > **Restart Instance**, and check whether the instances restart successfully.

- If yes, go to **Step 4**.
- If no, go to **Step 10**.

Step 4 Wait several minutes, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Increase the number of slots in each Supervisor.

Step 5 Log in to the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Storm** > **Configurations** > **All Configurations**.

Step 6 Increase the number of ports in the **supervisor.slots.ports** parameter of each Supervisor role and restart the instance.

Step 7 Wait several minutes, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Step 8 Perform capacity expansion for Supervisor.

Step 9 Wait several minutes, and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 10**.

NOTE

Services are interrupted when the Supervisor is being restarted. Then, services are restored after the restarting.

Collect fault information.

Step 10 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 11 Select **Storm** and **ZooKeeper** in the required cluster from the **Service** drop-down list box.

Step 12 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 13 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.183 ALM-26054 Nimbus Heap Memory Usage Exceeds the Threshold

Description

The system checks the heap memory usage of Storm Nimbus every 30 seconds and compares the actual usage with the threshold. The alarm is generated when the heap memory usage of Storm Nimbus exceeds the threshold (80% of the maximum memory by default) for 5 consecutive times.

Users can choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Storm** > **Nimbus** to change the threshold.

The alarm is cleared when the heap memory usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
26054	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

When the heap memory usage of Storm Nimbus is overhigh, frequent GCs occur. In addition, a memory overflow may occur so that the Yarn service is unavailable.

Possible Causes

The heap memory of the Storm Nimbus instance on the node is overused or the heap memory is inappropriately allocated. As a result, the usage exceeds the threshold.

Procedure

Check the heap memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms** > **Heap Memory Usage of Storm Nimbus Exceeds the Threshold** > **Location**. Check the host name of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Storm** > **Instance**. Click the instance for which the alarm is generated to go to the page for the instance. Click the drop-down menu in the chart area and choose **Customize** > **Nimbus** > **Heap Memory Usage of Nimbus**. Click **OK**.

Step 3 Check whether the used heap memory of Nimbus reaches the threshold (The default value is 80% of the maximum heap memory) specified for Nimbus.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Storm** > **Configurations** > **All Configurations** > **Nimbus** > **System**. Change the value of **-Xmx** in **NIMBUS_GC_OPTS** based on site requirements, and click **Save**. Click **OK**.

NOTE

- You are advised to set **-Xms** and **-Xmx** to the same value to prevent adverse impact on performance when JVM dynamically adjusts the heap memory size.
- The number of Workers grows as the Storm cluster scale increases. You can increase the value of **GC_OPTS** for Nimbus. The recommended value is as follows: If the number of Workers is 20, set **-Xmx** to a value greater than or equal to 1 GB. If the number of Workers exceeds 100, set **-Xmx** to a value greater than or equal to 5 GB.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 7 Select the following node in the required cluster from the **Service** drop-down list.

- NodeAgent
- Storm

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.184 ALM-27001 DBService Service Unavailable

Description

The alarm module checks the DBService service status every 30 seconds. This alarm is generated when the system detects that DBService service is unavailable.

This alarm is cleared when DBService service recovers.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
27001	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The database service is unavailable and cannot provide data import and query functions for upper-layer services, which results in some services exceptions.

Possible Causes

- The floating IP address does not exist.
- There is no active DBServer instance.
- The active and standby DBServer processes are abnormal.

Procedure

Check whether the floating IP address exists in the cluster environment.

Step 1 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **DBService** > **Instance**.

Step 2 Check whether the active instance exists.

- If yes, go to **Step 3**.
- If no, go to **Step 9**.

Step 3 Select the active DBServer instance and record the IP address.

Step 4 Log in to the host that corresponds to the preceding IP address as user **root**, and run the **ifconfig** command to check whether the DBService floating IP address exists on the node.

- If yes, go to **Step 5**.

- If no, go to **Step 9**.

Step 5 Run the **ping** *floatip* command to check whether the DBService floating IP address can be pinged successfully.

- If yes, go to **Step 6**.
- If no, go to **Step 9**.

Step 6 Log in to the host that corresponds to the DBService floating IP address as user **root**, and run the command to delete the floating IP address.

ifconfig *interface* **down**

Step 7 On the FusionInsight Manager home page, choose **Cluster >** *Name of the desired cluster* > **Services** > **DBService** > **More** > **Restart Service** to restart DBService, and check whether DBService is restarted successfully.

- If yes, go to **Step 8**.
- If no, go to **Step 9**.

Step 8 Wait for about 2 minutes and check whether the alarm is cleared in the alarm list.

- If yes, no further action is required.
- If no, go to **Step 14**.

Check the status of the active DBServer instance.

Step 9 Select the DBServer instance whose role status is abnormal and record the IP address.

Step 10 On the **Alarm** page, check whether **Process Fault** occurs in the DBServer instance on the host that corresponds to the IP address.

- If yes, go to **Step 11**.
- If no, go to **Step 14**.

Step 11 Handle the alarm according to "ALM-12007 Process Fault".

Step 12 Wait for about 5 minutes and check whether the alarm is cleared in the alarm list.

- If yes, no further action is required.
- If no, go to **Step 19**.

Check the status of the active and standby DBServers.

Step 13 Log in to the host that corresponds to the preceding IP address as user **root**, and run the **su - omm** command to switch to user **omm**.

Step 14 Run the **cd ${DBSERVER_HOME}** command to go to the installation directory of the DBService.

Step 15 Run the **sh sbin/status-dbserver.sh** command to view the status of the active and standby HA processes of DBService. Determine whether the status can be viewed successfully.

```
HAMode
double

NodeName              HostName            HAVersion            StartTime               HAActive
HAAllResOK          HARunPhase
10_5_89_12             host01            V100R001C01          2019-06-13 21:33:09      active
normal              Actived
```

```
10_5_89_66          host03          V100R001C01          2019-06-13 21:33:09     standby
normal          Deactived

NodeName              ResName           ResStatus           ResHAStatus          ResType
10_5_89_12          floatip          Normal              Normal              Single_active
10_5_89_12          gaussDB            Active_normal          Normal              Active_standby
10_5_89_66          floatip          Stopped             Normal              Single_active
10_5_89_66          gaussDB            Standby_normal          Normal              Active_standby
```

- If yes, go to **Step 16**.
- If no, go to **Step 19**.

Step 16 Check whether the active and standby HA processes are in the abnormal state.

- If yes, go to **Step 17**.
- If no, go to **Step 19**.

Step 17 On FusionInsight Manager, choose **Cluster >** *Name of the desired cluster* > **Services** > **DBService** > **More** > **Restart Service** to restart DBService, and check whether the system displays a message indicating that the restart is successful.

- If yes, go to **Step 18**.
- If no, go to **Step 19**.

Step 18 Wait for about 2 minutes and check whether the alarm is cleared in the alarm list.

- If yes, no further action is required.
- If no, go to **Step 19**.

Collect fault information.

Step 19 On FusionInsight Manager, choose **O&M** > **Log > Download**.

Step 20 Select **DBService** in the required cluster and **NodeAgent** from the **Service**.

Step 21 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 22 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.185 ALM-27003 DBService Heartbeat Interruption Between the Active and Standby Nodes

Description

This alarm is generated when the active or standby DBService node does not receive heartbeat messages from the peer node for 7 seconds.

This alarm is cleared when the heartbeat recovers.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
27003	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Local DBService HA Name	Specifies a local DBService HA.
Peer DBService HA Name	Specifies a peer DBService HA.

Impact on the System

During the DBService heartbeat interruption, only one node can provide the service. If this node is faulty, no standby node is available for failover and the service is unavailable.

Possible Causes

The link between the active and standby DBService nodes is abnormal.

Procedure

Check whether the network between the active DBService server and the standby DBService server is normal.

Step 1 In the alarm list on FusionInsight Manager, click ⌄ in the row where the alarm is located in the real-time alarm list and view the standby DBService server address.

Step 2 Log in to the active DBService server as user **root**.

Step 3 Run the **ping** *standby DBService heartbeat IP address* command to check whether the standby DBService server is reachable.

- If yes, go to **Step 6**.

- If no, go to **Step 4**.

Step 4 Contact the network administrator to check whether the network is faulty.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 Rectify the network fault and check whether the alarm is cleared from the alarm list.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 7 Select the following nodes in the required cluster from the **Service**:

- DBService
- Controller
- NodeAgent

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.186 ALM-27004 Data Inconsistency Between Active and Standby DBServices

Description

The system checks the data synchronization status between the active and standby DBService every 10 seconds. This alarm is generated when the synchronization status cannot be queried for six consecutive times or when the synchronization status is abnormal.

This alarm is cleared when the synchronization status becomes normal.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
27004	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Local DBService HA Name	Specifies the HA name of the local DBService.
Peer DBService HA Name	Specifies the HA name of the peer DBService.
SYNC_PERCENT	Specifies the synchronization percentage.

Impact on the System

When data is not synchronized between the active and standby DBServices, data may be lost or abnormal if the active instance becomes abnormal.

Possible Causes

- The network between the active and standby nodes is unstable.
- The standby DBService is abnormal.
- The standby node disk space is full.
- The CPU usage of the GaussDB process on the active DBService node is high. You need to locate the failure cause based on logs.

Procedure

Check whether the network between the active and standby nodes is normal.

Step 1 On FusionInsight Manager, choose **Cluster > Services > DBService > Instance**, check the service IP address of the standby DBServer instance.

Step 2 Log in to the active DBService node as user **root**.

Step 3 Run the **ping** *Standby DBService heartbeat IP address* command to check whether the standby DBService node is reachable.

- If yes, go to **Step 6**.
- If no, go to **Step 4**.

Step 4 Contact the network administrator to check whether the network is faulty.

- If yes, go to **Step 5**.
- If no, go to **Step 6**.

Step 5 Rectify the network fault and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Check whether the standby DBService is normal.

Step 6 Log in to the standby DBService node as user **root**.

Step 7 Run the **su - omm** command to switch to user **omm**.

Step 8 Go to the **${DBSERVER_HOME}/sbin** directory and run the **./status-dbserver.sh** command to check whether the GaussDB resource status of the standby DBService is normal. In the command output, check whether the following information is displayed in the row where **ResName** is **gaussDB**:

For example:

```
10_10_10_231 gaussDB Standby_normal Normal Active_standby
```

- If yes, go to **Step 9**.
- If no, go to **Step 16**.

Check whether the standby node disk space is full.

Step 9 Log in to the standby DBService node as user **root**.

Step 10 Run the **su - omm** command to switch to user **omm**.

Step 11 Go to the **${DBSERVER_HOME}** directory, and run the following commands to obtain the DBService data directory:

cd ${DBSERVER_HOME}

source .dbservice_profile

echo ${DBSERVICE_DATA_DIR}

Step 12 Run the **df -h** command to view the system disk partition usage information.

Step 13 Check whether the DBService data directory space is full.

- If yes, go to **Step 14**.
- If no, go to **Step 16**.

Step 14 Expand the disk capacity.

Step 15 After the disk capacity is expanded, wait 2 minutes and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 16**.

Collect fault information.

Step 16 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 17 In the **Service** area, select **DBService** of the target cluster and **OS**, **OS Statistics**, and **OS Performance** under **OMS**, and click **OK**.

Step 18 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 19 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.187 ALM-27005 Database Connections Usage Exceeds the Threshold

Description

The system checks the usage of the number of database connections of the nodes where DBServer instances are located every 30 seconds and compares the usage with the threshold. If the usage exceeds the threshold for five consecutive times (this number is configurable, and 5 is the default value), the system generates this alarm. The default usage threshold is 90%, and you can configure it based on site requirements.

The trigger count is configurable. This alarm is cleared in the following scenarios:

- The trigger count is 1, and the usage of the number of database connections is less than or equal to the threshold.
- The trigger count is greater than 1, and the usage of the number of database connections is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
27005	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

Upper-layer services may fail to connect to the DBService database, affecting services.

Possible Causes

- Too many database connections are used.
- The maximum number of database connections is improperly configured.
- The alarm threshold or alarm trigger count is improperly configured.

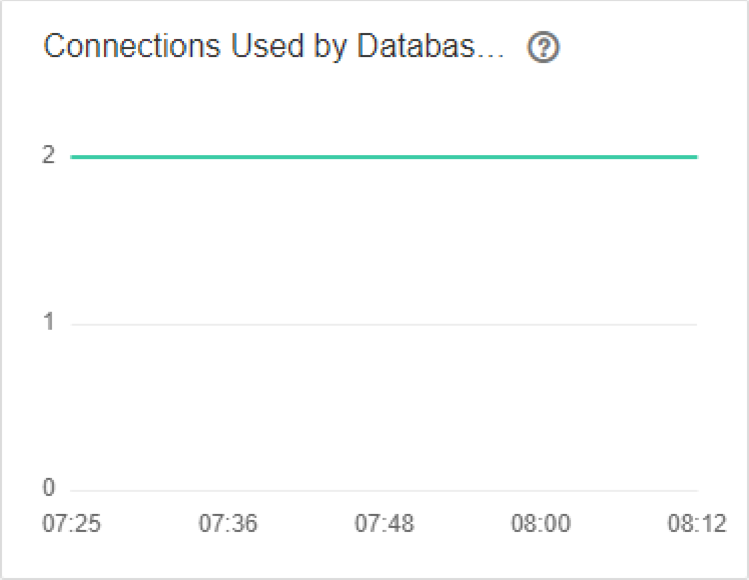
Procedure

Checking whether too many data connections are used

Step 1 On FusionInsight Manager, click DBService in the service list on the left navigation pane. The DBService monitoring page is displayed.

Step 2 Observe the number of connections used by the database user, as shown in **Figure 9-31**. Based on the service scenario, reduce the number of database user connections.

Figure 9-31 Number of connections used by database users



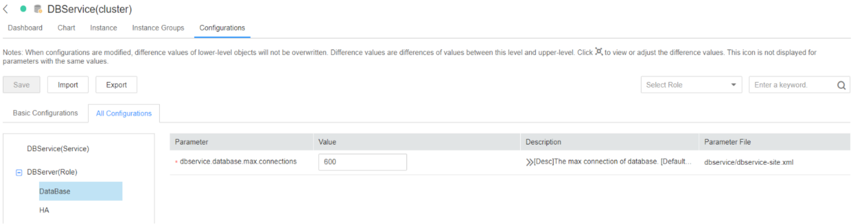
Step 3 Wait for 2 minutes and check whether the alarm is automatically cleared.

- If it is, no further action is required.
- If it is not, go to **Step 4**.

Checking whether the maximum number of database connections is properly configured

Step 4 Log in to FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **DBService** > **Configurations**. On the displayed page, select the **All Configurations** tab, and increase the maximum number of database connections based on service requirements, as shown in **Figure 9-32**. Click **Save**. In the displayed **Save configuration** dialog box, click **OK**.

Figure 9-32 Setting the maximum number of database connections



Notes: When configurations are modified, difference values of lower-level objects will not be overwritten. Difference values are differences of values between this level and upper-level. Click to view or adjust the difference values. This icon is not displayed for parameters with the same values.

Parameter	Value	Description	Parameter File
* dbservice.database.max.connections	600	[Desc]The max connection of database. [Default...	dbservice/dbservice-site.xml

Step 5 After the maximum number of database connections is changed, restart DBService (do not restart the upper-layer services).

Procedure: Log in to FusionInsight Manager and choose **Cluster** > *Name of the desired cluster* > **Services** > **DBService**. On the displayed page, choose **More** > **Restart Service**. Enter the password of the current login user and click **OK**. Do not select **Restart upper-layer services.**, click **OK**.

Step 6 After the service is restarted, wait for 2 minutes and check whether the alarm is cleared.

- If it is, no further action is required.

- If it is not, go to **Step 7**.

Checking whether the alarm threshold or trigger count is properly configured

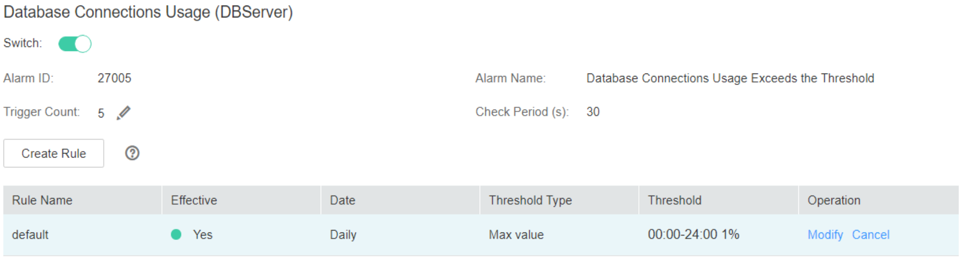
Step 7 Log in to FusionInsight Manager and change the alarm threshold and alarm trigger count based on the actual database connection usage.

Choose **O&M**> **Alarm** > **Thresholds** > *Name of the desired cluster* > **DBService > Database > Database Connections Usage (DBServer)**. In the **Database Connections Usage (DBServer)** area, click the pencil icon next to **Trigger Count**. In the displayed dialog box, change the trigger count, as shown in **Figure 9-33**.

NOTE

Trigger Count: If the usage of the number of database connections exceeds the threshold consecutively for more than the value of this parameter, an alarm is generated.

Figure 9-33 Setting alarm trigger count



Database Connections Usage (DBServer)

Switch: ●

Alarm ID: 27005

Alarm Name: Database Connections Usage Exceeds the Threshold

Trigger Count: 5

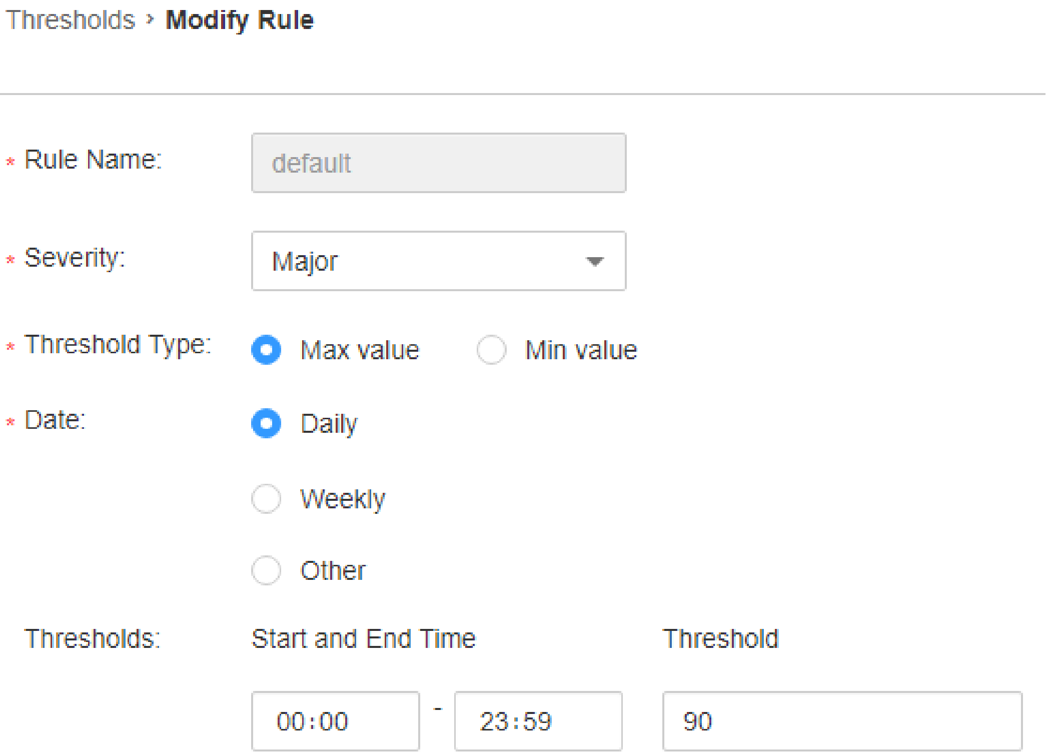
Check Period (s): 30

Create Rule

Rule Name	Effective	Date	Threshold Type	Threshold	Operation
default	● Yes	Daily	Max value	00:00-24:00 1%	Modify Cancel

Based on the actual database connection usage, choose **O&M** >**Alarm** > **Thresholds** > *Name of the desired cluster* > **DBService > Database > Database Connections Usage (DBServer)**. In the **Database Connections Usage (DBServer)** area, click **Modify** in the **Operation** column. In the **Modify Rule** dialog box, modify the required parameters and click **OK** as shown in **Figure 9-34**.

Figure 9-34 Set alarm threshold



Thresholds › **Modify Rule**

* Rule Name: default

* Severity: Major

* Threshold Type: ● Max value ○ Min value

* Date: ● Daily ○ Weekly ○ Other

Thresholds:

Start and End Time		Threshold
00:00	23:59	90

Step 8 Wait for 2 minutes and check whether the alarm is automatically cleared.

- If it is, no further action is required.
- If it is not, go to **Step 9**.

Collect fault information

Step 9 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 10 Select **DBService** in the required cluster from the **Service**.

Step 11 Specify the host for collecting logs by setting the **Host** parameter that is optional. By default, all hosts are selected.

Step 12 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 13 Contact the O&M personnel and send the collected fault logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.188 ALM-27006 Disk Space Usage of the Data Directory Exceeds the Threshold

Description

The system checks the disk space usage of the data directory on the active DBServer node every 30 seconds and compares the disk usage with the threshold. The alarm is generated when the disk space usage exceeds the threshold for five consecutive times (the default value). The number of consecutive times is configurable. The disk space usage threshold of the data directory is set to 80% by default, which is configurable as well.

The value of **hit number** is configurable. When the value is set to **1** and the disk space usage is lower than or equal to the threshold, the alarm is cleared. When the value is greater than 1 and the disk space usage is lower than 90% of the threshold, the alarm is cleared.

Attribute

Alarm ID	Alarm Severity	Auto Clear
27006	Major	Yes

Parameters

Name	Meaning
ClusterName	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
PartitionName	Specifies the disk partition where the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the actual indicator value exceeds this threshold, the alarm is generated.

Impact on the System

- Service processes become unavailable.

- When the disk space usage of the data directory exceeds 90%, the database reports the "Database Enters the Read-Only Mode" alarm and enters the read-only mode, which may cause service data loss.

Possible Causes

- The alarm threshold is improperly configured.
- The data volume of the database is too large or the disk configuration cannot meet service requirements, causing excessive disk usage.

Procedure

Check whether the threshold is set properly.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **DBService** > **Database** > **Disk Space Usage of the Data Directory** to check whether the alarm threshold is proper (the default value 80% is a proper value).

- If yes, go to **Step 3**.
- If no, go to **Step 2**.

Step 2 Change the alarm threshold based on the actual service situation.

Step 3 Choose **Cluster** > *Name of the desired cluster* > **Services** > **DBService**. On the **Dashboard** page, view the **Disk Space Usage of the Data Directory** chart and check whether the disk space usage of the data directory is lower than the threshold.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 Wait 2 minutes and check whether the alarm is automatically cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check whether large files are incorrectly written into the disk.

Step 5 Log in to the active DBService node as user **omm**.

Step 6 Run the following commands to view the files whose size exceeds 500 MB in the data directory and check whether there are large files incorrectly written into the directory:

```
source $DBSERVER_HOME/.dbservice_profile
```

```
find "$DBSERVICE_DATA_DIR"/../ -type f -size +500M
```

- If yes, go to **Step 7**.
- If no, go to **Step 8**.

Step 7 Handle the large files based on the actual scenario and check whether the alarm is cleared 2 minutes later.

- If yes, no further action is required.
- If no, go to **Step 8**.

Collect fault information.

Step 8 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 9 Expand the **Service** drop-down list, and select **DBService** for the target cluster.

Step 10 Specify the host for collecting logs by setting the **Host** parameter which is optional. By default, all hosts are selected.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.189 ALM-27007 Database Enters the Read-Only Mode

Description

The system checks the disk space usage of the data directory on the active DBServer node every 30 seconds. The alarm is generated when the disk space usage exceeds 90%.

The alarm is cleared when the disk space usage is lower than 80%.

Attribute

Alarm ID	Alarm Severity	Auto Clear
27007	Critical	Yes

Parameters

Name	Meaning
ClusterName	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.

Name	Meaning
Trigger Condition	Specifies the threshold triggering the alarm. If the actual indicator value exceeds this threshold, the alarm is generated.

Impact on the System

The database enters the read-only mode, causing service data loss.

Possible Causes

The disk configuration cannot meet service requirements. The disk usage reaches the upper limit.

Procedure

Check whether the disk space usage reaches the upper limit.

Step 1 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **DBService**.

Step 2 On the **Dashboard** page, view the **Disk Space Usage of the Data Directory** chart and check whether the disk space usage of the data directory exceeds 90%.

- If yes, go to **Step 3**.
- If no, go to **Step 13**.

Step 3 Log in to the active management node of the DBServer as user **omm** and run the following commands to check whether the database enters the read-only mode:

source $DBSERVER_HOME/.dbservice_profile

gsql -U omm -W *password* **-d postgres -p 20051**

show default_transaction_read_only;

NOTE

In the preceding commands, *password* indicates the password of user **omm** of the DBService database. You can run the **\q** command to exit the database.

Check whether the value of **default_transaction_read_only** is **on**.

```
POSTGRES=# show default_transaction_read_only;
 default_transaction_read_only
-------------------------------
 on
(1 row)
```

- If yes, go to **Step 4**.
- If no, go to **Step 13**.

Step 4 Run the following commands to open the **dbservice.properties** file:

source $DBSERVER_HOME/.dbservice_profile

```
vi ${DBSERVICE_SOFTWARE_DIR}/tools/dbservice.properties
```

Step 5 Change the value of **gaussdb_readonly_auto** to **OFF**.

Step 6 Run the following command to open the **postgresql.conf** file:

```
vi ${DBSERVICE_DATA_DIR}/postgresql.conf
```

Step 7 Delete **default_transaction_read_only = on**.

Step 8 Run the following command for the configuration to take effect:

```
gs_ctl reload -D ${DBSERVICE_DATA_DIR}
```

Step 9 Log in to FusionInsight Manager and choose **O&M** > **Alarm** > **Alarms**. On the right of the alarm "Database Enters the Read-Only Mode", click **Clear** in the **Operation** column. In the dialog box that is displayed, click **OK** to manually clear the alarm.

Step 10 Log in to the active management node of the DBServer as user **omm** and run the following commands to view the files whose size exceeds 500 MB in the data directory and check whether there are large files incorrectly written into the directory:

```
source $DBSERVER_HOME/.dbservice_profile
```

```
find "$DBSERVICE_DATA_DIR"/../ -type f -size +500M
```

- If yes, go to **Step 11**.
- If no, go to **Step 13**.

Step 11 Handle the files that are incorrectly written into the directory based on the actual scenario.

Step 12 Log in to FusionInsight Manager and choose **Cluster** > *Name of the desired cluster* > **Services** > **DBService**. On the **Dashboard** page, view the **Disk Space Usage of the Data Directory** chart and check whether the disk space usage is lower than 80%.

- If yes, no further action is required.
- If no, go to **Step 13**.

Collect fault information.

Step 13 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 14 Expand the **Service** drop-down list, and select **DBService** for the target cluster.

Step 15 Specify the host for collecting logs by setting the **Host** parameter which is optional. By default, all hosts are selected.

Step 16 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 17 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.190 ALM-38000 Kafka Service Unavailable

Description

The system checks the Kafka service status every 30 seconds. This alarm is generated when the Kafka service is unavailable.

This alarm is cleared when the Kafka service recovers.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
38000	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The cluster cannot provide the Kafka service, and users cannot perform new Kafka tasks.

Possible Causes

- The KrbServer service is abnormal.(Skip this step if the normal mode is used.)
- The ZooKeeper service is abnormal or does not respond.
- The Broker instance in the Kafka cluster are abnormal.

Procedure

Check the status of the KrbServer service. (Skip this step if the normal mode is used.)

Step 1 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services > KrbServer**.

Step 2 Check whether the running status of the KrbServer service is **Normal**.

- If yes, go to **Step 5**.
- If no, go to **Step 3**.

Step 3 Rectify the fault by following the steps provided in **ALM-25500 KrbServer Service Unavailable**.

Step 4 Perform **Step 2** again.

Check the status of the ZooKeeper cluster.

Step 5 Check whether the running status of the ZooKeeper service is **Normal**.

- If yes, go to **Step 8**.
- If no, go to **Step 6**.

Step 6 If ZooKeeper service is stopped, start it, else rectify the fault by following the steps provided in **ALM-13000 ZooKeeper Service Unavailable**.

Step 7 Perform **Step 5** again.

Check the Broker status.

Step 8 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Instance** to go to the Kafka instances page.

Step 9 Check whether all instances in **Roles** are running properly.

- If yes, go to **Step 11**.
- If no, go to **Step 10**.

Step 10 Select all Broker instances, choose **More** > **Restart Instance**, and check whether the instances restart successfully.

- If yes, go to **Step 11**.
- If no, go to **Step 13**.

Step 11 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** to check whether the running status is **Normal**.

- If yes, go to **Step 12**.
- If no, go to **Step 13**.

Step 12 Wait for 30 seconds and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 13**.

Collecting Fault Information

Step 13 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 14 Select **Kafka** in the required cluster from the **Service** drop-down list.

Step 15 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 16 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.191 ALM-38001 Insufficient Kafka Disk Capacity

Description

The system checks the Kafka disk usage every 60 seconds and compares the actual disk usage with the threshold. The disk usage has a default threshold. This alarm is generated when the disk usage is greater than the threshold.

You can change the threshold in **O&M** > **Alarm** > **Thresholds**. Under the service list, choose **Kafka > Disk > Broker Disk Usage (Broker)** and change the threshold.

When the **Trigger Count** is 1, this alarm is cleared when the Kafka disk usage is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the Kafka disk usage is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
38001	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.

Name	Meaning
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
PartitionName	Specifies the disk partition where the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

Kafka data write operations are affected.

Possible Causes

- The configuration (such as number and size) of the disks for storing Kafka data cannot meet the requirement of the current service traffic, due to which the disk usage reaches the upper limit.
- Data retention time is too long, due to which the data disk usage reaches the upper limit.
- The service plan does not distribute data evenly, due to which the usage of some disks reaches the upper limit.

Procedure

Check the disk configuration of Kafka data.

Step 1 On the FusionInsight Manager portal and click **O&M** > **Alarm** > **Alarms**.

Step 2 In the alarm list, locate the alarm and obtain **HostName** from **Location**.

Step 3 Click **Cluster** > *Name of the desired cluster* > **Hosts**.

Step 4 In the host list, click the host name obtained in **Step 2**.

Step 5 Check whether the **Disk** area contains the partition name in the alarm.

- If yes, go to **Step 6**.
- If no, manually clear the alarm and no further operation is required.

Step 6 Check whether the disk partition usage contained in the alarm reaches 100% in the **Disk** area.

- If yes, handle the alarm by following the instructions in **Related Information**.
- If no, go to **Step 7**.

Check the Kafka data storage duration.

Step 7 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Configurations**.

Step 8 Check whether the value of parameter **disk.adapter.enable** is set to **true**.

- If yes, go to **Step 10**.
- If no, go to **Step 9**.

Step 9 Set the value of **disk.adapter.enable** to **true**. Check whether the value of **adapter.topic.min.retention.hours** is properly set.

- If yes, go to **Step 10**.
- If no, adjust the data retention period based on service requirements.

NOTICE

If the disk auto-adaptation function is enabled, some historical data of specified topics is deleted. If the retention period of some topics cannot be adjusted, click **All Configurations** and add the topics to the value of the **disk.adapter.topic.blacklist** parameter.

Step 10 Wait 10 minutes and check whether the usage of faulty disks reduces.

- If yes, wait until the alarm is cleared.
- If no, go to **Step 11**.

Check the Kafka data plan.

Step 11 In the **Instance** area, click **Broker**. In the **Real Time** area of Broker, Click the drop-down menu in the Chart area and choose **Customize** to customize monitoring items.

Step 12 In the dialog box, select **Disk** > **Broker Disk Usage** and click **OK**.

The Kafka disk usage information is displayed.

Step 13 View the information in **Step 12** to check whether there is only the disk parathion for which the alarm is generated in **Step 2**.

- If yes, go to **Step 14**.
- If no, go to **Step 15**.

Step 14 Perform disk planning and mount a new disk again. Go to the **Instance Configurations** page of the node for which the alarm is generated, modify **log.dirs**, add other disk directories, and restart the Kafka instance.

Step 15 Determine whether to shorten the data retention time configured on Kafka based on service requirements and service traffic.

- If yes, go to **Step 16**.
- If no, go to **Step 17**.

Step 16 Log in to FusionInsight Manager, select **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Configurations**, and click **All Configurations**. In the search box on the right, enter **log.retention.hours**. The value of the parameter indicates the default data retention time of the topic. You can change the value to a smaller one.

NOTE

- For a topic whose data retention time is configured alone, the modification of the data retention time on the Kafka Service Configuration page does not take effect.
- To modify the data retention time for a topic, use the Kafka client command-line interface (CLI) to configure the topic.

  Example: **kafka-topics.sh --zookeeper "***ZooKeeper IP address***:2181/kafka" --alter --topic "***Topic bane***" --config retention.ms= "***retention time***"**

Step 17 Check whether the usage of some disks reaches the upper limit due to unreasonable configuration of the partitions of some topics. For example, the number of partitions configured for a topic with large data volume is smaller than the number of disks. In this case, the data is not evenly allocated to disks.

NOTE

If you do not know which topic has large data volume, you can log in to an instance node based on the host node information obtained in **Step 2**, and go to the data directory (directory specified by **log.dirs** before the modification in **Step 14**) to check whether there is topic with partition that use large disk space.

- If yes, go to **Step 18**.
- If no, go to **Step 19**.

Step 18 In the Kafka client CLI, run the following command to perform partition capacity expansion for the topic:

kafka-topics.sh --zookeeper "*ZooKeeper IP address***:2181/kafka" --alter --topic "***Topic name***" --partitions="***New number of partitions***"**

NOTE

- You are advised to set the new number of partitions to a multiple of the number of Kafka data disks.
- The step may not quickly clear the alarm, and you need to modify the data retention time in **Step 11** to gradually balance data allocation.

Step 19 Determine whether to perform capacity expansion.

NOTE

You are advised to perform capacity expansion for Kafka when the current disk usage exceeds 80%.

- If yes, go to **Step 20**.
- If no, go to **Step 21**.

Step 20 Expand the disk capacity and check whether the alarm is cleared after capacity expansion.

- If yes, no further action is required.
- If no, go to **Step 22**.

Step 21 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 22**.

Collect fault information.

Step 22 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 23 Select **Kafka** in the required cluster from the **Service** drop-down list.

Step 24 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 25 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

Step 1 Log in to FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Instance**, stop the Broker instance whose status is **Restoring**, record the management IP address of the node where the Broker instance is located, and record **broker.id**. The value can be obtained by using the following method: Click the role name. On the **Configurations** page, select **All Configurations**, and search for the **broker.id** parameter.

Step 2 Log in to the recorded management IP address as user **root**, and run the **df -lh** command to view the mounted directory whose disk usage is 100%, for example, **${BIGDATA_DATA_HOME}/kafka/data1**.

Step 3 Go to the directory, run the **du -sh *** command to view the size of each file in the directory,check whether files other than **kafka-logs** exist, and determine whether these files can be deleted or migrated.

- If yes, go to **Step 8**.
- If no, go to **Step 4**.

Step 4 Go to the **kafka-logs** directory, run the **du -sh *** command, select a partition folder to be moved. The naming rule is **Topic name-Partition ID**. Record the topic and partition.

Step 5 Modify the **recovery-point-offset-checkpoint** and **replication-offset-checkpoint** files in the **kafka-logs** directory in the same way.

1. Decrease the number in the second line in the file. (To remove multiple directories, the number deducted is equal to the number of files to be removed.)
2. Delete the line of the to-be-removed partition. (The line structure is "Topic name Partition ID Offset". Save the data before deletion. Subsequently, the content must be added to the file of the same name in the destination directory.)

Step 6 Modify the **recovery-point-offset-checkpoint** and **replication-offset-checkpoint** files in the destination data directory. For example, **${BIGDATA_DATA_HOME}/kafka/data2/kafka-logs** in the same way.

- Increase the number in the second line in the file. (To move multiple directories, the number added is equal to the number of files to be moved.)

- Add the to-be moved partition to the end of the file. (The line structure is "Topic name Partition ID Offset". You can copy the line data saved in **Step 5**.)

Step 7 Move the partition to the destination directory. After the partition is moved, run the **chown omm:wheel -R** *Partition directory* command to modify the directory owner group for the partition.

Step 8 Log in to FusionInsight Manager and choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Instance** to start the Broker instance.

Step 9 Wait for 5 to 10 minutes and check whether the health status of the Broker instance is **Normal**.

- If yes, resolve the disk capacity insufficiency problem according to the handling method of "ALM-38001 Insufficient Kafka Disk Space" after the alarm is cleared.
- If no, contact the O&M personnel.

----End

9.13.192 ALM-38002 Kafka Heap Memory Usage Exceeds the Threshold

Description

The system checks the Kafka service status every 30 seconds. The alarm is generated when the heap memory usage of a Kafka instance exceeds the threshold (95% of the maximum memory) for 10 consecutive times.

When the **Trigger Count** is 1, this alarm is cleared when the heap memory usage is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the heap memory usage is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
38002	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.

Name	Meaning
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available Kafka heap memory is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The heap memory of the Kafka instance is overused or the heap memory is inappropriately allocated.

Procedure

Check heap memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms** > **Kafka Heap Memory Usage Exceeds the Threshold** > **Location**. Check the host name of the instance involved in this alarm.

Step 2 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Instance**. Click the instance for which the alarm is generated to go to the page for the instance. Click the drop-down list in the upper right corner of the chart area, choose **Customize** > **Process** > **Heap Memory Usage of Kafka**, and click **OK**.

Step 3 Check whether the used heap memory of Kafka reaches 95% of the maximum heap memory specified for Kafka.

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Check the heap memory size configured for Kafka.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Configurations** > **All Configurations**> **Broker(Role)** > **Environment**. Increase the value of **KAFKA_HEAP_OPTS** by referring to the Note.

NOTE

- It is recommended that **-Xmx** and **-Xms** be set to the same value.
- You are advised to view **Heap Memory Usage of Kafka** by referring to **Step 2**, and set the value of **KAFKA_HEAP_OPTS** to twice the value of **Heap Memory Used by Kafka.**

Step 5 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 7 Select **Kafka** in the required cluster from the **Service** drop-down list.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.193 ALM-38004 Kafka Direct Memory Usage Exceeds the Threshold

Description

The system checks the direct memory usage of the Kafka service every 30 seconds. This alarm is generated when the direct memory usage of a Kafka instance exceeds the threshold (80% of the maximum memory) for 10 consecutive times.

When the **Trigger Count** is 1, this alarm is cleared when the direct memory usage is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the direct memory usage is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
38004	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available direct memory of the Kafka service is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The direct memory of the Kafka instance is overused or the direct memory is inappropriately allocated.

Procedure

Check the direct memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms** > **Kafka Direct Memory Usage Exceeds the Threshold** > **Location** to check the host name of the instance for which the alarm is generated.

Step 2 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Instance**. Click the instance for which the alarm is generated to go to the page for the instance. Click the drop-down menu in the Chart area and choose **Customize** > **Process** > **Kafka Direct Memory Usage**, and click **OK**.

Step 3 Check whether the used direct memory of Kafka reaches 80% of the maximum direct memory specified for Kafka.

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Check the direct memory size configured for the Kafka.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Configurations** > **All Configurations** > **Broker(Role)** > **Environment** to increase the value of **-Xmx** configured in the **KAFKA_HEAP_OPTS** parameter by referring to the Note.

NOTE

- It is recommended that **-Xmx** and **-Xms** be set to the same value.
- You are advised to view **Kafka Direct Memory Usage** by referring to **Step 2**, and set the value of **KAFKA_HEAP_OPTS** to twice the value of **Direct Memory Used by Kafka.**

Step 5 Save the configuration and restart the Kafka service.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 8 Select **Kafka** in the required cluster from the **Service** drop-down list.

Step 9 Click in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.194 ALM-38005 GC Duration of the Broker Process Exceeds the Threshold

Description

The system checks the garbage collection (GC) duration of the Broker process every 60 seconds. This alarm is generated when the GC duration exceeds the threshold (12 seconds by default) for 3 consecutive times.

When the **Trigger Count** is 1, this alarm is cleared when the GC duration is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the GC duration is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
38005	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

A long GC duration of the Broker process may interrupt the services.

Possible Causes

The Kafka GC duration of the node is too long or the heap memory is inappropriately allocated. As a result, GCs occur frequently.

Procedure

Check the GC duration.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms** > **GC Duration of the Broker Process Exceeds the Threshold** > **Location**. Check the host name of the instance involved in this alarm.

Step 2 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Instance**. Click the instance for which the alarm is generated to go to the page for the instance. Click the drop-down list in the upper right corner of the chart area, choose **Customize** > **Process** > **Broker GC Duration per Minute**, and click **OK**.

Step 3 Check whether the GC duration of the Broker process collected every minute exceeds the threshold (12 seconds by default).

- If yes, go to **Step 4**.
- If no, go to **Step 7**.

Check the direct memory size configured for the Kafka.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Configurations** > **All Configurations** > **Broker(Role)**

> **Environment** to increase the value of **-Xmx** configured in the **KAFKA_HEAP_OPTS** parameter by referring to the Note.

NOTE

- It is recommended that **-Xmx** and **-Xms** be set to the same value.
- You are advised to set the value of **KAFKA_HEAP_OPTS** to twice the value of **Direct Memory Used by Kafka.**

  On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Instance**. Click the instance for which the alarm is generated to go to the page for the instance. Click the drop-down list in the upper right corner of the chart area and choose **Customize** > **Process** > **Kafka Direct Memory Resource Status** to check the value of **Direct Memory Used by Kafka**.

Step 5 Save the configuration and restart the Kafka service.

Step 6 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 8 Select **Kafka** in the required cluster from the **Service** drop-down list.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.195 ALM-38006 Percentage of Kafka Partitions That Are Not Completely Synchronized Exceeds the Threshold

Description

The system checks the percentage of Kafka partitions that are not completely synchronized to the total number of partitions every 60 seconds. This alarm is generated when the percentage exceeds the threshold (50% by default) for 3 consecutive times.

When the **Trigger Count** is 1, this alarm is cleared when the percentage is less than or equal to the threshold. When the **Trigger Count** is greater than 1, this alarm is cleared when the percentage is less than or equal to 90% of the threshold.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
38006	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

Too many Kafka partitions that are not completely synchronized affect service reliability. In addition, data may be lost when leaders are switched.

Possible Causes

Some nodes where the Broker instance resides are abnormal or stop running. As a result, replicas of some partitions in Kafka are out of the in-sync replicas (ISR) set.

Procedure

Check Broker instances.

Step 1 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Instance**. The Kafka instances page is displayed.

Step 2 Check whether faulty nodes exist among all Broker nodes.

- If yes, record the host name of the node and go to **Step 3**.
- If no, go to **Step 5**.

Step 3 On the FusionInsight Manager portal, click **O&M** > **Alarm** > **Alarms** to check whether the fault described in **Step 2** exists in the alarm information and handle the alarm based on corresponding methods.

Step 4 On the FusionInsight Manager portal, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Instance**. The Kafka instances page is displayed.

Step 5 Check whether stopped nodes exist among all Broker instance.

- If yes, go to **Step 6**.
- If no, go to **Step 7**.

Step 6 Select all stopped Broker instances and click **Start Instance**.

Step 7 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Collect fault information.

Step 8 On the FusionInsight Manager portal, choose **O&M** > **Log** > **Download**.

Step 9 Select **Kafka** in the required cluster from the **Service** drop-down list.

Step 10 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.196 ALM-38007 Status of Kafka Default User Is Abnormal

Description

The system checks the default user of Kafka every 60 seconds. This alarm is generated when the system detects that the user status is abnormal.

Trigger Count is set to **1**. This alarm is cleared when the user status becomes normal.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
38007	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host name for which the alarm is generated.
Trigger Condition	Specifies the condition that the Kafka default user status is abnormal.

Impact on the System

If the Kafka default user status is abnormal, metadata synchronization between Brokers and interaction between Kafka and ZooKeeper will be affected, affecting service production, consumption, and topic creation and deletion.

Possible Causes

- The Sssd service is abnormal.
- Some Broker instances stop running.

Procedure

Check whether the Sssd service is abnormal.

Step 1 On the FusionInsight Manager portal, choose **O&M** > **Alarm** > **Alarms** > **Status of Kafka Default User Is Abnormal** > **Location** to check the host name of the instance for which the alarm is generated.

Step 2 Find the host information in the alarm information and log in to the host.

Step 3 Run the **id -Gn kafka** command and check whether "No such user" is displayed in the command output.

- If yes, record the host name of the node and go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On the FusionInsight Manager home page, choose **O&M** > **Alarm** > **Alarms**. Check whether there is **Sssd Service Exception** in the alarm information. If there is, handle the alarm based on alarm information.

Check the running status of the Broker instance.

Step 5 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Instance**. The Kafka instance page is displayed.

Step 6 Check whether there are stopped nodes on all Broker instances.

- If yes, go to **Step 7**.
- If no, go to **Step 8**.

Step 7 Select all stopped Broker instances and click **Start Instance**.

Step 8 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect fault information.

Step 9 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 10 In the **Service** area, select **Kafka** in the required cluster.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.197 ALM-38008 Abnormal Kafka Data Directory Status

Description

The system checks the Kafka data directory status every 60 seconds. This alarm is generated when the system detects that the status of a data directory is abnormal.

Trigger Count is set to **1**. This alarm is cleared when the data directory status becomes normal.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
38008	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host name for which the alarm is generated.
DirName	Specifies the directory name for which the alarm is generated.
Trigger Condition	Specifies the condition that the Kafka data directory status is abnormal.

Impact on the System

If the Kafka data directory status is abnormal, the current replicas of all partitions in the data directory are brought offline, and the data directory status of multiple nodes is abnormal at the same time. As a result, some partitions may become unavailable.

Possible Causes

- The data directory permission is tampered with.
- The disk where the data directory is located is faulty.

Procedure

Check the permission on the faulty data directory.

Step 1 Find the host information in the alarm information and log in to the host.

Step 2 In the alarm information, check whether the data directory and its subdirectories belong to the omm:wheel group.

- If yes, record the host name of the node and go to **Step 4**.
- If no, go to **Step 3**.

Step 3 Restore the owner group of the data directory and its subdirectories to omm:wheel.

- If yes, go to **Step 6**.
- If no, go to **Step 5**.

Check whether the disk where the data directory is located is faulty.

Step 4 In the upper-level directory of the data directory, create and delete files as user **omm**. Check whether data read/write on the disk is normal.

Step 5 Replace or repair the disk where the data directory is located to ensure that data read/write on the disk is normal.

Step 6 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Instance**. On the Kafka instance page that is displayed, restart the Broker instance on the host recorded in **Step 2**.

Step 7 After Broker is started, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 8**.

Collect fault information.

Step 8 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 9 In the **Service** area, select **Kafka** in the required cluster.

Step 10 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.198 ALM-38009 Busy Broker Disk I/Os (Applicable to Versions Later Than MRS 3.1.0)

NOTE

This section applies to versions later than MRS 3.1.0.

Description

The system checks the I/O status of each Kafka disk every 60 seconds. This alarm is generated when the disk I/O of a Kafka data directory on a broker exceeds the threshold (80% by default).

Its **Trigger Count** is **3**. This alarm is cleared when the disk I/O is lower than the threshold (80% by default).

Attribute

Alarm ID	Alarm Severity	Auto Clear
38009	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
DataDirectoryName	Specifies the name of the Kafka data directory with frequent disk I/Os.

Impact on the System

The disk partition has frequent I/Os. Data may fail to be written to the Kafka topic for which the alarm is generated.

Possible Causes

- There are many replicas configured for the topic.
- The parameter for batch writing producer's messages is inappropriately configured. The service traffic of this topic is too heavy, and the current partition configuration is inappropriate.

Procedure

Check the number of topic replicas.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. Locate the row that contains this alarm, click ⌄, and view the host name in **Location**.

Step 2 On FusionInsight Manager, choose **Cluster**, click the name of the desired cluster, choose **Services** > **Kafka** > **KafkaTopic Monitor**, search for the topic for which the alarm is generated, and check the number of replicas.

Step 3 Reduce the replication factors of the topic (for example, reduce to **3**) if the number of replicas is greater than 3.

Run the following command on the FusionInsight client to replan the replicas of Kafka topics:

```
kafka-reassign-partitions.sh --zookeeper {zk_host}:{port}/kafka --reassignment-json-file {manual assignment json file path} --execute
```

For example:

```
/opt/client/Kafka/kafka/bin/kafka-reassign-partitions.sh --zookeeper 10.149.0.90:2181,10.149.0.91:2181,10.149.0.92:2181/kafka --reassignment-json-file expand-cluster-reassignment.json --execute
```

NOTE

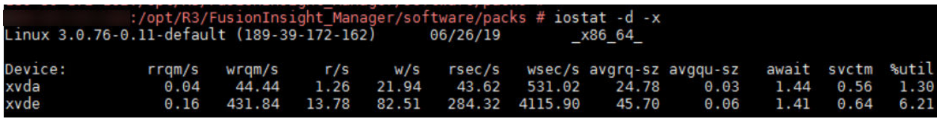
In the **expand-cluster-reassignment.json** file, describe the brokers to which the partitions of the topic are migrated in the following format: {"partitions":[{"topic": "*topicName*","partition": 1,"replicas": [1,2,3] }],"version":1}

Step 4 Observe for a period of time and check whether the alarm is cleared. If the alarm persists, go to **Step 5**.

Check the partition planning of the topic.

Step 5 On the **KafkaTopic Monitor** page, view **Topic Input Traffic** in the **Topic Traffic** area of each topic, obtain the topic with the largest value, and check the partitions of this topic as well as information about the host of these partitions.

Step 6 Log in to the host queried in **Step 5** and run the **iostat -d -x** command to check the **%util** value of each disk.



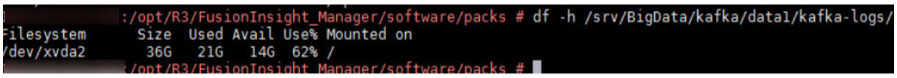
```
[redacted]:/opt/R3/FusionInsight_Manager/software/packs # iostat -d -x
Linux 3.0.76-0.11-default (189-39-172-162)      06/26/19        _x86_64_

Device:         rrqm/s   wrqm/s     r/s     w/s   rsec/s   wsec/s avgrq-sz avgqu-sz   await  svctm  %util
xvda              0.04    44.44    1.26   21.94    43.62   531.02    24.78     0.03    1.44   0.56   1.30
xvde              0.16   431.84   13.78   82.51   284.32  4115.90    45.70     0.06    1.41   0.64   6.21
```

- If the **%util** value of each disk exceeds the threshold (**80%** by default), expand the Kafka disk capacity. After the capacity expansion, replan the topic partitions by referring to **Step 3**.
- If the **%util** values of the disks vary greatly, check the disk partition configuration of Kafka. For example, check the value of **log.dirs** in the **${BIGDATA_HOME}/FusionInsight_HD_8.1.0.1/1_14_Broker/etc/server.properties** file.

  Run the following command to view the **Filesystem** information:

  df -h *log.dirs value*

  The command output is as follows.

```
[redacted]:/opt/R3/FusionInsight_Manager/software/packs # df -h /srv/BigData/kafka/data1/kafka-logs/
Filesystem      Size  Used Avail Use% Mounted on
/dev/xvda2       36G   21G   14G  62% /
[redacted]:/opt/R3/FusionInsight_Manager/software/packs #
```

- If the partition where Filesystem is located matches the partition with a high **%util** value, plan Kafka partitions on idle disks, configure **log.dirs** as an idle disk directory, and replan topic partitions by referring to **Step 3**. Ensure that the partitions of the topic are evenly distributed to each disk.

Step 7 Observe for a period of time and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, repeat **Step 5** to **Step 6** three times. Then, go to **Step 8**.

Step 8 Observe for a period of time and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect fault information.

Step 9 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 10 Expand the **Service** drop-down list, and select **Kafka** for the target cluster.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.199 ALM-38010 Topics with Single Replica

Description

The system checks the number of replicas of each topic every 60 seconds on the node where the Kafka Controller resides. This alarm is generated when there is one replica for a topic.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
38010	Warning	No

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
TopicName	Specifies the list of topics for which the alarm is generated.

Impact on the System

There is the single point of failure (SPOF) risk for topics with only one replica. When the node where the replica resides becomes abnormal, the partition does not have a leader, and services on the topic are affected.

Possible Causes

- The number of replicas for the topic is incorrectly configured.

Procedure

Check the number of replicas for the topic.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**, click ⌄ of this alarm, and view the **TopicName** list in **Location**.

Step 2 Check whether replicas need to be added for the topic for which the alarm is generated.

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 On the FusionInsight client, re-plan topic replicas and describe the partition distribution of the topic in the **add-replicas-reassignment.json** file in the following format: {"partitions":[{"topic": "*topic name*","partition": 1,"replicas": [1,2] }],"version":1}. Then, run the following command to add replicas:

kafka-reassign-partitions.sh --zookeeper *{zk_host}:{port}***/kafka --reassignment-json-file** *{manual assignment json file path}* **--execute**

For example:

/opt/client/Kafka/kafka/bin/kafka-reassign-partitions.sh --zookeeper 192.168.0.90:2181,192.168.0.91:2181,192.168.0.92:2181/kafka --reassignment-json-file add-replicas-reassignment.json --execute

Step 4 Run the following command to check the task execution progress:

kafka-reassign-partitions.sh --zookeeper *{zk_host}:{port}***/kafka --reassignment-json-file** *{manual assignment json file path}* **--verify**

For example:

/opt/client/Kafka/kafka/bin/kafka-reassign-partitions.sh --zookeeper 192.168.0.90:2181,192.168.0.91:2181,192.168.0.92:2181/kafka --reassignment-json-file add-replicas-reassignment.json --verify

Step 5 After completing the handling operations or confirming that the alarm has no impact, manually clear the alarm on FusionInsight Manager.

Step 6 After a period of time, check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 7**.

Collect fault information.

Step 7 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 8 In the **Service** area, select **Kafka** in the required cluster.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

If the alarm has no impact, manually clear the alarm.

Related Information

None

9.13.200 ALM-38011 User Connection Usage on Broker Exceeds the Threshold

Description

The system checks the number of connections of each user on Broker every 30 seconds. This alarm is generated when the connection usage of a user on the Broker exceeds the threshold (80% by default) for 5 consecutive times.

The number of times that smoothing is performed is **5**. This alarm is cleared when the connection usage of a user on the Broker is less than the threshold.

The alarm can be automatically cleared. However, if the number of connections of a user suddenly becomes **0** and no connection is created, the alarm cannot be automatically cleared. You need to manually clear it.

Attribute

Alarm ID	Alarm Severity	Automatically Cleared
38011	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
UserName	Specifies the username for which the alarm is generated.

Impact on the System

If the number of connections of a user is excessive, the user cannot create new connections to the Broker.

Possible Causes

- The number of connections (created by a user) used by the client exceeds the preset threshold.
- The threshold for the connection usage does not meet service requirements.

Procedure

Check the number of connections established by the same user on the client.

Step 1 On the FusionInsight Manager home page, choose **O&M** > **Alarm** > **Alarms** > **User Connection Usage on Broker Exceeds the Threshold**. Check the host name and username of the Broker instance for which the alarm is generated in **Location**.

Step 2 On FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Instance**. Click the instance for which the alarm is generated to go to the page for the instance. Click the drop-down list in the upper right corner of the chart area, choose **Customize > Other**, and select **User Connection Usage on Broker**, **Maximum Number of User Connections on Broker**, and **Number of User Connections on Broker** to view the number of the current user connections on the Broker.

Step 3 Observe the number of real-time connections of the current alarm user and check whether the real-time monitoring data of the current user exists.

- If yes, go to **Step 4**.
- If no, the current user has disconnected all connections. You need to clear the alarm manually, and no further action is required.

NOTE

After the alarm user disconnects all connections, the monitoring data of the user disappears. In this case, the alarm will not be automatically cleared. You need to manually clear it.

Step 4 Check whether the user is authorized by the service side.

If yes, go to **Step 7**.

If no, go to **Step 5**.

Step 5 Run the following command on the client to limit the number of connections of the user. There are two configuration rules based on the following commands:

1. For the specific Broker and user, run the following command:

   kafka-configs.sh --bootstrap-server *<broker ip:port>* **--alter --add-config 'max.connections.per.user.overrides=[***<username>***:***<connection.number>***]' --entity-type brokers --entity-name** *<broker.id>* **--command-config Kafka/kafka/config/producer.properties**

NOTE

For unauthorized users, confirm with the service side to reduce the maximum number of connections of an unauthorized user or set the maximum number of connections to **0**.

In the command, you need to specify the IP address and port number of Broker, set values of configuration items, and specify the **brokerId** and **username**. Here, the user refers to the authorized Kerberos user.

The configuration updated using the command line tool can take effect dynamically. The configuration becomes invalid after the service is restarted. To make the configuration take effect after the restart, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Configurations** > **All Configurations**> **Broker** > **Server** on the FusionInsight Manager home page and update the configuration to **max.connections.per.user.overrides**.

2. For the specific use and default Broker (that is, all Broker instances in the cluster), run the following command:

   kafka-configs.sh --bootstrap-server *<broker ip:port>* **--alter --add-config 'max.connections.per.user.overrides=[***<username>***:***<connection.number>***]' --entity-type brokers ---entity-default --command-config Kafka/kafka/config/client.properties**

   Example:

   kafka-configs.sh --bootstrap-server 10.153.3.26:21007 --alter --add-config 'max.connections.per.user.overrides=[showcase:4]' --entity-type brokers --entity-name 1 --command-config Kafka/kafka/config/client.properties

Step 6 Check whether the maximum number of connections is **0** and whether the number of connections of the current user decreases or remains unchanged according to **Step 2**.

- If yes, manually clear the alarm and no further action is required.
- If no, go to **Step 7**.

Step 7 Check whether the number of real-time connections and connection usage of the current user are sharply increased when they are compared with historical data, and whether have exceeded the specified maximum number of connections.

- If yes, go to **Step 8**.
- If no, go to **Step 9**.

NOTE

If there is an obvious increase after the comparison and the maximum number of connections has reached the preset value, the connections of the user may be abnormal. You need to confirm with the service party.

Check whether the number of user connections meets service requirements.

Step 8 Check whether the number of connections of the user meets service requirements.

- If yes, go to **Step 9**.
- If no, contact the service party to rectify the fault.

NOTE

If the number of user connections is abnormal, contact the service party to rectify the fault from the following aspects:

- Check whether new services are added so that the number of user connections increases sharply.
- Check whether handle leakage occurs on the code at the service side.

Step 9 Consider whether to increase the maximum number of connections of the user.

- If yes, go to **Step 10**.
- If no, go to **Step 12**.

Step 10 Increase the maximum number of connections based on the service requirements. Set the number of connections of the user on the Kafka client. For details, see **Step 5**.

Step 11 Wait for several minutes and then check whether the alarm is automatically cleared.

- If yes, go to **Step 12**.
- If no, go to **Step 2**.

Step 12 Determine whether to add the user to the whitelist based on service requirements on the service side.

- If yes, go to **Step 13**.
- If no, go to **Step 15**.

NOTE

To add a user to the whitelist, you need to restart the Kafka service. However, this operation will cause service interruption and affect service running. Therefore, you must confirm with the service side before performing this operation.

Step 13 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **Kafka** > **Configurations** > **All Configurations** > **Broker(Role)** > **Server** to add the user to the **max.connections.per.user.whitelist** configuration item.

Step 14 Restart the service for the modification to take effect. In addition, you need to manually clear the alarm, and no further action is required.

Collect the fault information.

Step 15 On the FusionInsight Manager homepage, choose **O&M** > **Log** > **Download**.

Step 16 Expand the **Service** drop-down list, and select **Kafka** for the target cluster.

Step 17 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 18 Contact the O&M personnel and send the collected fault logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.201 ALM-43001 Spark2x Service Unavailable

Description

The system checks the Spark2x service status every 300 seconds. This alarm is generated when the Spark2x service is unavailable.

This alarm is cleared when the Spark2x service recovers.

Attribute

Alarm ID	Alarm Severity	Auto Clear
43001	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The Spark tasks submitted by users fail to be executed.

Possible Causes

- The KrbServer service is abnormal.
- The LdapServer service is abnormal.
- ZooKeeper is abnormal.
- HDFS is abnormal.
- Yarn is abnormal.
- The corresponding Hive service is abnormal.
- The Spark2x assembly package is abnormal.

Procedure

If the alarm is abnormal Spark2x assembly packet, the Spark packet is abnormal. Wait for about 10 minutes. The alarm is automatically cleared.

Check whether service unavailability alarms exist in services that Spark2x depends on.

Step 1 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Alarm** > **Alarms**.

Step 2 Check whether the following alarms exist in the alarm list:

- ALM-25500 KrbServer Service Unavailable
- ALM-25000 LdapServer Service Unavailable
- ALM-13000 ZooKeeper Service Unavailable
- ALM-14000 HDFS Service Unavailable
- ALM-18000 Yarn Service Unavailable
- ALM-16004 Hive Service Unavailable

NOTE

If the multi-instance function is enabled for the cluster and multiple Spark2x services are installed, check the Spark2x service for which the alarm is generated based on the value of **ServiceName** in location information and check whether the Hive service is faulty. Spark2x corresponds to Hive, spark2x1 corresponds to Hive1, and other services follow the same rule.

- If yes, go to **Step 3**.
- If no, go to **Step 4**.

Step 3 Handle the alarms based on the troubleshooting methods provided in the alarm help.

After the alarm is cleared, wait a few minutes and check whether the alarm GuardianService Unavailable is cleared.

- If yes, no further action is required.
- If no, go to **Step 4**.

Collect fault information.

Step 4 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 5 In the **Service** area, select the following nodes of the desired cluster. (Hive is the specific Hive service determined based on **ServiceName** in the alarm location information).

- KrbServer
- LdapServer
- ZooKeeper
- HDFS
- Yarn
- Hive

Step 6 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 7 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.202 ALM-43006 Heap Memory Usage of the JobHistory2x Process Exceeds the Threshold

Description

The system checks the JobHistory2x Process status every 30 seconds. The alarm is generated when the heap memory usage of a JobHistory2x Process exceeds the threshold (95% of the maximum memory).

Attribute

Alarm ID	Alarm Severity	Auto Clear
43006	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available JobHistory2x Process heap memory is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The heap memory of the JobHistory2x Process is overused or the heap memory is inappropriately allocated.

Procedure

Check heap memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** and select the alarm whose **ID** is **43006**. Check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Instance** and click the JobHistory2x for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the Chart area and choose **Customize** > **Memory > JobHistory2x Memory Usage Statistics** from the drop-down list box in the upper right corner and click **OK**. Check whether the used heap memory of the JobHistory2x Process reaches the threshold(default value is 95%) of the maximum heap memory specified for JobHistory2x.

- If yes, go to **Step 3**.
- If no, go to **Step 7**.

Step 3 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. Click **JobHistory2x** by which the alarm is reported to go to the **Dashboard** page, click the drop-down list in the upper right corner of the chart area, choose **Customize** > **Memory > Statistics for the heap memory of the JobHistory2x Process**, and click **OK**. Based on the alarm generation time, check the values of the used heap memory of the JobHistory2x process in the corresponding period and obtain the maximum value.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Configurations**, and click **All Configurations**. Choose **JobHistory2x** > **Default**. The default value of **SPARK_DAEMON_MEMORY** is 4GB. You can change the value according to the following rules: Ratio of the maximum heap memory usage of the JobHistory2x to the **Threshold** of the **JobHistory2x Heap Memory Usage Statistics (JobHistory2x)** in the alarm period. If this alarm is generated occasionally after the parameter value is adjusted, increase the value by 0.5 times. If the alarm is frequently reported after the parameter value is adjusted, increase the value by 1 time.

NOTE

On the FusionInsight Manager home page, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* **>** **Spark2x** > **Memory** >**JobHistory2x Heap Memory Usage Statistics (JobHistory2x)** to view **Threshold**.

Step 5 Restart all JobHistory2x instances.

Step 6 After 10 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **Spark2x** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.203 ALM-43007 Non-Heap Memory Usage of the JobHistory2x Process Exceeds the Threshold

Description

The system checks the JobHistory2x Process status every 30 seconds. The alarm is generated when the non-heap memory usage of a JobHistory2x Process exceeds the threshold (95% of the maximum memory).

Attribute

Alarm ID	Alarm Severity	Auto Clear
43007	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.

Name	Meaning
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available JobHistory2x Process non-heap memory is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The non-heap memory of the JobHistory2x Process is overused or the non-heap memory is inappropriately allocated.

Procedure

Check non-heap memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** and select the alarm whose **ID** is **43007**. Check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Instance** and click the JobHistory2x for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the Chart area and choose **Customize** > **Memory** > **JobHistory2x Memory Usage Statistics** from the drop-down list box in the upper right corner and click **OK**, Check whether the used non-heap memory of the JobHistory2x Process reaches the threshold(default value is 95%) of the maximum non-heap memory specified for JobHistory2x.

- If yes, go to **Step 3**.
- If no, go to **Step 7**.

Step 3 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. Click **JobHistory2x** by which the alarm is reported to go to the **Dashboard** page, click the drop-down list in the upper right corner of the chart area, choose **Customize** > **Memory > Statistics for the non-heap memory of the JobHistory2x Process**, and click **OK**. Based on the alarm generation time, check the values of the used non-heap memory of the JobHistory2x process in the corresponding period and obtain the maximum value.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Configurations**, and click **All Configurations**. Choose **JobHistory2x** > **Default**. You can change the value of **-**

XX:MaxMetaspaceSize in **SPARK_DAEMON_JAVA_OPTS** according to the following rules: Ratio of the JobHistory2x non-heap memory usage to the **Threshold** of **JobHistory2x Non-Heap Memory Usage Statistics (JobHistory2x)** in the alarm period.

NOTE

On the FusionInsight Manager home page, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* **> Spark2x** > **Memory** >**JobHistory2x Non-Heap Memory Usage Statistics (JobHistory2x)** to view **Threshold**.

Step 5 Restart all JobHistory2x instances.

Step 6 After 10 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **Spark2x** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.204 ALM-43008 The Direct Memory Usage of the JobHistory2x Process Exceeds the Threshold

Description

The system checks the JobHistory2x Process status every 30 seconds. The alarm is generated when the direct memory usage of a JobHistory2x Process exceeds the threshold (95% of the maximum memory).

Attribute

Alarm ID	Alarm Severity	Auto Clear
43008	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available JobHistory2x Process direct memory is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The direct memory of the JobHistory2x Process is overused or the direct memory is inappropriately allocated.

Procedure

Check direct memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** and select the alarm whose **ID** is **43008**. Check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Instance** and click the JobHistory2x for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the Chart area and choose **Customize** > **Memory** > **JobHistory2x Memory Usage Statistics** from the drop-down list box in the upper right corner and click **OK**. Check whether the used direct memory of the JobHistory2x Process reaches the threshold(default value is 95%) of the maximum direct memory specified for JobHistory2x.

- If yes, go to **Step 3**.
- If no, go to **Step 7**.

Step 3 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. Click **JobHistory2x** by which the alarm is reported to go to the **Dashboard** page, click the drop-down list in the upper right corner of the chart area, choose **Customize** > **Memory > Direct Memory of JobHistory2x**, and click **OK**. Based on the alarm generation time, check the values of the used direc memory of the JobHistory2x process in the corresponding period and obtain the maximum value.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Configurations**, and click **All Configurations**. Choose **JobHistory2x** > **Default**. The default value of **-XX:MaxDirectMemorySize** in **SPARK_DAEMON_JAVA_OPTS** is 512 MB. You can change the value according to the following rules: Ratio of the maximum direct memory usage of the JobHistory2x to the **Threshold** of the **JobHistory2x Direct Memory Usage Statistics (JobHistory2x)** in the alarm period. If this alarm is generated occasionally after the parameter value is adjusted, increase the value by 0.5 times. If the alarm is frequently reported after the parameter value is adjusted, increase the value by 1 time. It is recommended that the value be less than or equal to the value of **SPARK_DAEMON_MEMORY**.

NOTE

On the FusionInsight Manager home page, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* **> Spark2x** > **Memory** >**JobHistory2x Direct Memory Usage Statistics (JobHistory2x)** to view **Threshold**.

Step 5 Restart all JobHistory2x instances.

Step 6 After 10 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **Spark2x** in the required cluster from the **Service**.

Step 9 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.205 ALM-43009 JobHistory2x Process GC Time Exceeds the Threshold

Description

The system checks the garbage collection (GC) time of the JobHistory2x Process every 60 seconds. This alarm is generated when the detected GC time exceeds the threshold (exceeds 5 seconds for three consecutive checks.) To change the threshold, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Spark2x** > **GC Time** > **Total GC time in milliseconds (JobHistory2x)**. This alarm is cleared when the JobHistory2x GC time is shorter than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
43009	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the GC time exceeds the threshold, JobHistory2x maybe run in low performance.

Possible Causes

The memory of JobHistory2x is overused, the heap memory is inappropriately allocated. As a result, GCs occur frequently.

Procedure

Check the GC time.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** and select the alarm whose **ID** is **43009**. Check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Instance** and click the JobHistory2x for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the Chart area and choose **Customize** > **GC Time** > **Garbage Collection (GC) Time of JobHistory2x** from the drop-down list box in the upper right corner and click **OK** to check whether the GC time is longer than the threshold(default value: 12 seconds).

- If yes, go to **Step 3**.
- If no, go to **Step 6**.

Step 3 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Configurations**, and click **All Configurations**. Choose **JobHistory2x** > **Default**. The default value of **SPARK_DAEMON_MEMORY** is 4GB. You can change the value according to the following rules: If this alarm is generated occasionally, increase the value by 0.5 times. If the alarm is frequently reported, increase the value by 1 time.

Step 4 Restart all JobHistory2x instances.

Step 5 After 10 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager interface of active and standby clusters, choose **O&M** > **Log > Download**.

Step 7 Select **Spark2x** in the required cluster from the **Service**.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.206 ALM-43010 Heap Memory Usage of the JDBCServer2x Process Exceeds the Threshold

Description

The system checks the JDBCServer2x Process status every 30 seconds. The alarm is generated when the heap memory usage of a JDBCServer2x Process exceeds the threshold (95% of the maximum memory).

Attribute

Alarm ID	Alarm Severity	Auto Clear
43010	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available JDBCServer2x Process heap memory is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The heap memory of the JDBCServer2x Process is overused or the heap memory is inappropriately allocated.

Procedure

Check heap memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** and select the alarm whose **ID** is **43010**. Check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Instance** and click the JDBCServer2x for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the Chart area and choose **Customize** > **Memory** > **JDBCServer2x Memory Usage Statistics** from the drop-down list box in the upper right corner and click **OK**. Check whether the used heap memory of the JDBCServer2x Process reaches the threshold(default value is 95%) of the maximum heap memory specified for JDBCServer2x.

- If yes, go to **Step 3**.
- If no, go to **Step 7**.

Step 3 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. Click **JDBCServer2x** by which the alarm is reported to go to the **Dashboard** page, click the drop-down list in the upper right corner of the chart area, choose **Customize** > **Memory > Statistics for the heap memory of the JDBCServer2x Process**, and click **OK**. Based on the alarm generation time, check the values of the used heap memory of the JDBCServer2x process in the corresponding period and obtain the maximum value.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Configurations**, and click **All Configurations**. Choose **JDBCServer2x** > **Tuning**. The default value of **SPARK_DRIVER_MEMORY** is 4 GB. You can change the value according to the following rules: Ratio of the maximum heap memory usage of the JobHistory2x to the **Threshold** of the **JDBCServer2x Heap Memory Usage Statistics (JDBCServer2x)** in the alarm period. If this alarm is generated occasionally after the parameter value is adjusted, increase the value by 0.5 times. If the alarm is frequently reported after the parameter value is adjusted, increase the value by 1 time. In the case of large service volume and high concurrency, add instances.

NOTE

On the FusionInsight Manager home page, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* **> Spark2x** > **Memory** >**JDBCServer2x Heap Memory Usage Statistics (JDBCServer2x)** to view **Threshold**.

Step 5 Restart all JDBCServer2x instances.

Step 6 After 10 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **Spark2x** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.207 ALM-43011 Non-Heap Memory Usage of the JDBCServer2x Process Exceeds the Threshold

Description

The system checks the JDBCServer2x Process status every 30 seconds. The alarm is generated when the non-heap memory usage of an JDBCServer2x Process exceeds the threshold (95% of the maximum memory).

Attribute

Alarm ID	Alarm Severity	Auto Clear
43011	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available JDBCServer2x Process non-heap memory is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The non-heap memory of the JDBCServer2x Process is overused or the non-heap memory is inappropriately allocated.

Procedure

Check non-heap memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** and select the alarm whose **ID** is **43011**. Check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Instance** and click the JDBCServer2x for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the Chart area and choose **Customize** > **Memory** > **JDBCServer2x Memory Usage Statistics** from the drop-down list box in the upper right corner and click **OK**, Check whether the used non-heap memory of the JDBCServer2x Process reaches the threshold(default value is 95%) of the maximum non-heap memory specified for JDBCServer2x.

- If yes, go to **Step 3**.
- If no, go to **Step 7**.

Step 3 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. Click **JDBCServer2x** by which the alarm is reported to go to the **Dashboard** page, click the drop-down list in the upper right corner of the chart area, choose **Customize** > **Memory > Statistics for the non-heap memory of the JDBCServer2x Process**, and click **OK**. Based on the alarm generation time, check the values of the used non-heap memory of the JDBCServer2x process in the corresponding period and obtain the maximum value.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Configurations**, and click **All Configurations**. Choose **JDBCServer2x** > **Tuning**. You can change the value of **-XX: MaxMetaspaceSize** in **spark.driver.extraJavaOptions** according to the following rules: Ratio of the JDBCServer2x non-heap memory usage to the **Threshold** of **JDBCServer2x Non-Heap Memory Usage Statistics (JDBCServer2x)** in the alarm period.

NOTE

On the FusionInsight Manager home page, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* **> Spark2x** > **Memory** > **JDBCServer2x Non-Heap Memory Usage Statistics (JDBCServer2x)** to view **Threshold**.

Step 5 Restart all JDBCServer2x instances.

Step 6 After 10 minutes, check whether the alarm is cleared.

- If yes, no further action is required.

- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log >Download**.

Step 8 Select **Spark2x** in the required cluster from the **Service**.

Step 9 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.208 ALM-43012 Direct Heap Memory Usage of the JDBCServer2x Process Exceeds the Threshold

Description

The system checks the JDBCServer2x Process status every 30 seconds. The alarm is generated when the direct heap memory usage of a JDBCServer2x Process exceeds the threshold (95% of the maximum memory).

Attribute

Alarm ID	Alarm Severity	Auto Clear
43012	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.

Name	Meaning
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Specifies the threshold triggering the alarm. If the current indicator value exceeds this threshold, the alarm is generated.

Impact on the System

If the available JDBCServer2x Process direct heap memory is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The direct heap memory of the JDBCServer2x Process is overused or the direct heap memory is inappropriately allocated.

Procedure

Check direct heap memory usage.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** and select the alarm whose **ID** is **43012**. Check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Instance** and click the JDBCServer2x for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the Chart area and choose **Customize** > Memory > **JDBCServer2x Memory Usage Statistics** from the drop-down list box in the upper right corner and click **OK**. Check whether the used direct heap memory of the JDBCServer2x Process reaches the threshold(default value is 95%) of the maximum direct heap memory specified for JDBCServer2x.

- If yes, go to **Step 3**.
- If no, go to **Step 7**.

Step 3 On the FusionInsight Manager home page, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. Click **JDBCServer2x** by which the alarm is reported to go to the **Dashboard** page, click the drop-down list in the upper right corner of the chart area, choose **Customize** > **Memory > Direct Memory of JDBCServer2x**, and click **OK**. Based on the alarm generation time, check the values of the used direct memory of the JDBCServer2x process in the corresponding period and obtain the maximum value.

Step 4 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Configurations**, and click **All Configurations**. Choose **JDBCServer2x** > **Tuning**. The default value of **-XX:MaxDirectMemorySize** in **spark.driver.extraJavaOptions** is 512 MB. You can change the value according to the following rules: Ratio of the maximum direct memory usage of the

JDBCServer2x to the **Threshold** of the **JDBCServer2x Direct Memory Usage Statistics (JDBCServer2x)** in the alarm period. If this alarm is generated occasionally after the parameter value is adjusted, increase the value by 0.5 times. If the alarm is frequently reported after the parameter value is adjusted, increase the value by 1 time. In the case of large service volume and high service concurrency, you are advised to add instances.

NOTE

On the FusionInsight Manager home page, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* **> Spark2x** > **Memory** >**JDBCServer2x Direct Memory Usage Statistics (JDBCServer2x)** to view **Threshold**.

Step 5 Restart all JDBCServer2x instances.

Step 6 After 10 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 7**.

Collect fault information.

Step 7 On the FusionInsight Manager portal, choose **O&M** > **Log > Download**.

Step 8 Select **Spark2x** in the required cluster from the **Service**.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.209 ALM-43013 JDBCServer2x Process GC Time Exceeds the Threshold

Description

The system checks the garbage collection (GC) time of the JDBCServer2x Process every 60 seconds. This alarm is generated when the detected GC time exceeds the threshold (exceeds 5 seconds for three consecutive checks.) To change the threshold, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Spark2x** > **GC Time** > **Total GC time in milliseconds (JDBCServer2x)**. This alarm is cleared when the JDBCServer2x GC time is shorter than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
43013	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service name for which the alarm is generated.
RoleName	Specifies the role name for which the alarm is generated.
HostName	Specifies the object (host ID) for which the alarm is generated.
Trigger Condition	Generates an alarm when the actual indicator value exceeds the specified threshold.

Impact on the System

If the GC time exceeds the threshold, JDBCServer2x maybe run in low performance.

Possible Causes

The memory of JDBCServer2x is overused, the heap memory is inappropriately allocated. As a result, GCs occur frequently.

Procedure

Check the GC time.

Step 1 On the FusionInsight Manager portal, choose **O&M > Alarm > Alarms** and select the alarm whose **ID** is **43013**. Check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Instance** and click the JDBCServer2x for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the Chart area and choose **Customize** > **GC Time** > **Garbage Collection (GC) Time of JDBCServer2x** from the drop-down list box in the upper right corner and click **OK** to check whether the GC time is longer than the threshold(default value: 12 seconds).

- If yes, go to **Step 3**.
- If no, go to **Step 6**.

Step 3 On the FusionInsight Manager portal, choose **Cluster >** *Name of the desired cluster* **> Services** > **Spark2x** > **Configurations**, and click **All Configurations**. Choose **JDBCServer2x** > **Default**. The default value of **SPARK_DRIVER_MEMORY** is 4 GB. You can change the value according to the following rules: If this alarm is generated occasionally, increase the value by 0.5 times. If the alarm is frequently reported, increase the value by 1 time. In the case of large service volume and high service concurrency, you are advised to add instances.

Step 4 Restart all JDBCServer2x instances.

Step 5 After 10 minutes, check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect fault information.

Step 6 On the FusionInsight Manager interface of active and standby clusters, choose **O&M** > **Log > Download**.

Step 7 Select **Spark2x** in the required cluster from the **Service**.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Related Information

None

9.13.210 ALM-43017 JDBCServer2x Process Full GC Number Exceeds the Threshold

Description

The system checks the number of Full garbage collection (GC) times of the JDBCServer2x process every 60 seconds. This alarm is generated when the detected Full GC number exceeds the threshold (exceeds 12 for three consecutive checks.) You can change the threshold by choosing **O&M > Alarm** > **Thresholds** > *Name of the desired cluster* > **Spark2x** > **GC number** > **Full GC Number of JDBCServer2x**. This alarm is cleared when the Full GC number of the JDBCServer2x process is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
43017	Major	Yes

Parameters

Name	Description
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

The performance of the JDBCServer2x process is affected, or even the JDBCServer2x process is unavailable.

Possible Causes

The heap memory usage of the JDBCServer2x process is excessively large, or the heap memory is inappropriately allocated. As a result, Full GC occurs frequently.

Procedure

Check the number of Full GCs.

Step 1 Log in to FusionInsight Manager, choose **O&M** > **Alarm > Alarms**, select this alarm, and check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. On the displayed page, click the JDBCServer2x for which the alarm is reported. On the **Dashboard** page that is displayed, click the drop-down menu in the Chart area and choose **Customize** > **GC Number** > **Full GC Number of JDBCServer2x** in the upper right corner and click **OK**. Check whether the number of Full GCs of the JDBCServer2x process is greater than the threshold(default value: 12).

- If it is, go to **Step 3**.

- If it is not, go to **Step 6**.

Step 3 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Configurations** > **All Configurations**. On the displayed page, choose **JDBCServer2x** > **Tuning**. The default value of **SPARK_DRIVER_MEMORY** is 4GB. You can change the value according to the following rules: If this alarm is generated occasionally, increase the value by 0.5 times. If the alarm is frequently reported, increase the value by 1 time. In the case of large service volume and high concurrency, add instances.

Step 4 Restart all JDBCServer2x instances.

Step 5 After 10 minutes, check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 6**.

Collect fault information.

Step 6 Log in to FusionInsight Manager, and choose **O&M** > **Log** > **Download**.

Step 7 Select **Spark2x** in the required cluster from the **Service** drop-down list.

Step 8 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 10 minutes before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.211 ALM-43018 JobHistory2x Process Full GC Number Exceeds the Threshold

Description

The system checks the number of Full garbage collection (GC) times of the JobHistory2x process every 60 seconds. This alarm is generated when the detected Full GC number exceeds the threshold (exceeds 12 for three consecutive checks.) You can change the threshold by choosing **O&M** > **Alarm > Thresholds** > *Name of the desired cluster* > **Spark2x** > **GC number** > **Full GC Number of JobHistory2x**. This alarm is cleared when the Full GC number of the JobHistory2x process is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
43018	Major	Yes

Parameters

Name	Description
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

The performance of the JobHistory2x process is affected, or even the JobHistory2x process is unavailable.

Possible Causes

The heap memory usage of the JobHistory2x process is excessively large, or the heap memory is inappropriately allocated. As a result, Full GC occurs frequently.

Procedure

Check the number of Full GCs.

Step 1 Log in to FusionInsight Manager, choose **O&M** > **Alarm > Alarms**, select this alarm, and check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. On the displayed page, click the JobHistory2x for which the alarm is reported. On the **Dashboard** page that is displayed, click the drop-down menu in the Chart area and choose **Customize** > **GC Number** > **Full GC Number of JobHistory2x** in the upper right corner and click **OK**. Check whether the number of Full GCs of the JobHistory2x process is greater than the threshold(default value: 12).

- If it is, go to **Step 3**.

- If it is not, go to **Step 6**.

Step 3 Choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Configurations** > **All Configurations**. On the displayed page, choose **JobHistory2x** > **Default.** The default value of **SPARK_DAEMON_MEMORY** is 4GB. You can change the value according to the following rules: If this alarm is generated occasionally, increase the value by 0.5 times. If the alarm is frequently reported, increase the value by 1 time.

Step 4 Restart all JobHistory2x instances.

Step 5 After 10 minutes, check whether the alarm is cleared.

- If it is, no further action is required.
- If it is not, go to **Step 6**.

Collect fault information.

Step 6 Log in to FusionInsight Manager, and choose **O&M** > **Log** > **Download**.

Step 7 Select **Spark2x** in the required cluster from the **Service**.

Step 8 Click ✎ in the upper right corner. In the displayed dialog box, set **Start Date** and **End Date** to 10 minutes before and after the alarm generation time respectively and click **OK**. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected logs.

----End

Alarm Clearing

This alarm will be automatically cleared after the fault is rectified.

Related Information

None

9.13.212 ALM-43019 Heap Memory Usage of the IndexServer2x Process Exceeds the Threshold

Description

The system checks the IndexServer2x process status every 30 seconds. The alarm is generated when the heap memory usage of a IndexServer2x process exceeds the threshold (95% of the maximum memory).

Attribute

Alarm ID	Severity	Auto Clear
43019	Major	Yes

Parameters

Parameter	Description
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

If the available IndexServer2x process heap memory is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The heap memory of the IndexServer2x process is overused or the heap memory is inappropriately allocated.

Procedure

Check the heap memory usage.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm > Alarms**. In the displayed alarm list, choose the alarm for which the ID is **43019**, and check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. Click the IndexServer2x that reported the alarm to go to the **Dashboard** page. Click the drop-down list in the upper right corner of the chart area, and choose **Customize** > **Memory** > **IndexServer2x Memory Usage Statistics** > **OK**. Check whether the heap memory used by the IndexServer2x process reaches the maximum heap memory threshold (95% by default).

- If the threshold is reached, go to **Step 3**.
- If the threshold is not reached, go to **Step 7**.

Step 3 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. Click the IndexServer2x that reported the alarm to go to the **Dashboard** page. Click the drop-down list in the upper right corner of the chart area, and choose **Customize** > **Memory** > **Statistics for the heap memory of the IndexServer2x Process** > **OK**. Based on the alarm generation time, check the values of the used heap memory of the IndexServer2x process in the corresponding period and obtain the maximum value.

Step 4 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Configurations** > **All Configuration** > **IndexServer2x**> **Tuning**. The default value of the **SPARK_DRIVER_MEMORY** parameter is 4 GB. You can change the value based on the ratio of the maximum heap memory used by the IndexServer2x process to the threshold specified by **IndexServer2x Heap Memory Usage Statistics (IndexServer2x)** in the alarm period. If the alarm persists after the parameter value is changed, increase the value by 0.5 times. If the alarm is generated frequently, double the rate.

> **NOTE**
>
> On FusionInsight Manager, you can choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Spark2x** > **Memory** > **IndexServer2x Heap Memory Usage Statistics (IndexServer2x)** to view the threshold.

Step 5 Restart all IndexServer2x instances.

Step 6 After 10 minutes, check whether the alarm is cleared.

- If the alarm is cleared, no further action is required.
- If the alarm is not cleared, go to **Step 7**.

Collect fault information.

Step 7 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 8 Expand the **Service** drop-down list, and select **Spark2x** for the target cluster.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and provide the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Reference

None

9.13.213 ALM-43020 Non-Heap Memory Usage of the IndexServer2x Process Exceeds the Threshold

Description

The system checks the IndexServer2x process status every 30 seconds. The alarm is generated when the non-heap memory usage of the IndexServer2x process exceeds the threshold (95% of the maximum memory).

Attribute

Alarm ID	Severity	Auto Clear
43020	Major	Yes

Parameters

Parameter	Description
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

If the available IndexServer2x process non-heap memory is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The non-heap memory of the IndexServer2x process is overused or the non-heap memory is inappropriately allocated.

Procedure

Check non-heap memory usage.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm > Alarms**. In the displayed alarm list, choose the alarm for which the ID is **43020**, and check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. Click the IndexServer2x that reported the alarm to go to the **Dashboard** page. Click the drop-down list in the upper right corner of the chart area, and choose **Customize** > **Memory > IndexServer2x Memory Usage Statistics** > **OK**. Check whether the non-heap memory used by the IndexServer2x process reaches the maximum non-heap memory threshold (95% by default).

- If the threshold is reached, go to **Step 3**.

- If the threshold is not reached, go to **Step 7**.

Step 3 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. Click the IndexServer2x that reported the alarm to go to the **Dashboard** page. Click the drop-down list in the upper right corner of the chart area, and choose **Customize** > **Memory** > **Statistics for the non-heap memory of the IndexServer2x Process** > **OK**. Based on the alarm generation time, check the values of the used non-heap memory of the IndexServer2x process in the corresponding period and obtain the maximum value.

Step 4 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Configurations** > **All Configurations** > **IndexServer2x**> **Tuning**. You can change the value of **XX:MaxMetaspaceSize** in the **spark.driver.extraJavaOptions** parameter based on the ratio of the maximum non-heap memory used by the IndexServer2x process to the threshold specified by **IndexServer2x Non-Heap Memory Usage Statistics (IndexServer2x)** in the alarm period.

NOTE

On FusionInsight Manager, you can choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Spark2x** > **Memory** > **IndexServer2x Non-Heap Memory Usage Statistics (IndexServer2x)** to view the threshold.

Step 5 Restart all IndexServer2x instances.

Step 6 After 10 minutes, check whether the alarm is cleared.

- If the alarm is cleared, no further action is required.
- If the alarm is not cleared, go to **Step 7**.

Collect fault information.

Step 7 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 8 Expand the **Service** drop-down list, and select **Spark2x** for the target cluster.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and provide the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Reference

None

9.13.214 ALM-43021 Direct Memory Usage of the IndexServer2x Process Exceeds the Threshold

Description

The system checks the IndexServer2x process status every 30 seconds. The alarm is generated when the direct heap memory usage of a IndexServer2x process exceeds the threshold (95% of the maximum memory).

Attribute

Alarm ID	Severity	Auto Clear
43021	Major	Yes

Parameters

Parameter	Description
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

If the available IndexServer2x process direct memory is insufficient, a memory overflow occurs and the service breaks down.

Possible Causes

The direct heap memory of the IndexServer2x process is overused or the direct heap memory is inappropriately allocated.

Procedure

Check direct heap memory usage.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm > Alarms**. In the displayed alarm list, choose the alarm for which the ID is **43021**, and check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. Click the IndexServer2x that reported the alarm to go to the **Dashboard** page. Click the drop-down list in the upper right corner of the chart area, and choose **Customize** > **Memory** > **IndexServer2x Memory Usage Statistics** > **OK**. Check whether the direct memory used by the IndexServer2x process reaches the maximum direct memory threshold.

- If the threshold is reached, go to **Step 3**.
- If the threshold is not reached, go to **Step 7**.

Step 3 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance**. Click the IndexServer2x that reported the alarm to go to the **Dashboard** page. Click the drop-down list in the upper right corner of the chart area, and choose **Customize** > **Memory** > **Direct Memory of IndexServer2x** > **OK**. Based on the alarm generation time, check the values of the used direct memory of the IndexServer2x process in the corresponding period and obtain the maximum value.

Step 4 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Configurations** > **All Configurations** > **IndexServer2x**> **Tuning**. You can change the value of **XX:MaxDirectMemorySize** (the default value is 512 MB) in the **spark.driver.extraJavaOptions** parameter based on the ratio of the maximum direct memory used by the IndexServer2x process to the threshold specified by **IndexServer2x Direct Memory Usage Statistics (IndexServer2x)** in the alarm period. If the alarm persists after the parameter value is changed, increase the value by 0.5 times. If the alarm is generated frequently, double the rate.

NOTE

On FusionInsight Manager, you can choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Spark2x** > **Memory** > **IndexServer2x Direct Memory Usage Statistics (IndexServer2x)** to view the threshold.

Step 5 Restart all IndexServer2x instances.

Step 6 After 10 minutes, check whether the alarm is cleared.

- If the alarm is cleared, no further action is required.
- If the alarm is not cleared, go to **Step 7**.

Collect fault information.

Step 7 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 8 Expand the **Service** drop-down list, and select **Spark2x** for the target cluster.

Step 9 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 10 Contact the O&M personnel and provide the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Reference

None

9.13.215 ALM-43022 IndexServer2x Process GC Time Exceeds the Threshold

Description

The system checks the GC time of the IndexServer2x process every 60 seconds. This alarm is generated when the detected GC time exceeds the threshold (12 seconds) for three consecutive times. To change the threshold, choose **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Spark2x** > **GC Time** > **Total GC time in milliseconds (IndexServer2x)**. This alarm is cleared when the IndexServer2x GC time is shorter than or equal to the threshold.

Attribute

Alarm ID	Severity	Auto Clear
43022	Major	Yes

Parameters

Parameter	Description
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

If the GC time exceeds the threshold, IndexServer2x may run in low performance or even unavailable.

Possible Causes

The heap memory of the IndexServer2x process is overused or the heap memory is inappropriately allocated. As a result, GC occurs frequently.

Procedure

Check the GC time.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm > Alarms**. In the displayed alarm list, choose the alarm with ID **43022**, and check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance** and click the IndexServer2x for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the Chart area and choose **Customize** > **GC Time** > **Garbage Collection (GC) Time of IndexServer2x** from the drop-down list box in the upper right corner and click **OK** to check whether the GC time is longer than the threshold (default value: 12 seconds).

- If the threshold is reached, go to **Step 3**.
- If the threshold is not reached, go to **Step 6**.

Step 3 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Configurations** > **All Configurations** > **IndexServer2x** > **Default**. The default value of the **SPARK_DRIVER_MEMORY** is 4 GB. You can change the value according to the following rules: Increase the value of the **SPARK_DRIVER_MEMORY** parameter 1.5 times to its default value. If this alarm is still generated occasionally after the adjustment, increase the value by 0.5 times. Double the value if the alarm is reported frequently.

Step 4 Restart all IndexServer2x instances.

Step 5 After 10 minutes, check whether the alarm is cleared.

- If the alarm is cleared, no further action is required.
- If the alarm is not cleared, go to **Step 6**.

Collect fault information.

Step 6 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Spark2x** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and provide the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Reference

None

9.13.216 ALM-43023 IndexServer2x Process Full GC Number Exceeds the Threshold

Description

The system checks the Full GC number of the IndexServer2x process every 60 seconds. This alarm is generated when the detected Full GC number exceeds the threshold (12) for three consecutive times. You can change the threshold by choosing **O&M** > **Alarm** > **Thresholds** > *Name of the desired cluster* > **Spark2x** > **GC Number** > **Full GC Number of IndexServer2x**. This alarm is cleared when the Full GC number of the IndexServer2x process is less than or equal to the threshold. This alarm is cleared when the Full GC number of the IndexServer2x process is less than or equal to the threshold.

Attribute

Alarm ID	Severity	Auto Clear
43023	Major	Yes

Parameters

Parameter	Description
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

If the GC number exceeds the threshold, IndexServer2x maybe run in low performance or even unavailable.

Possible Causes

The heap memory of the IndexServer2x process is overused or the heap memory is inappropriately allocated. As a result, Full GC occurs frequently.

Procedure

Check the number of Full GCs.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm > Alarms**. In the displayed alarm list, choose the alarm with the ID **43023**, and check the **RoleName** in **Location** and confirm the IP address of **HostName**.

Step 2 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Instance** and click the IndexServer2x for which the alarm is generated to go to the **Dashboard** page. Click the drop-down menu in the chart area and choose **Customize** > **GC Number > Full GC Number of IndexServer2x** from the drop-down list box in the upper right corner and click **OK** to check whether the GC number is larger than the threshold (default value: 12).

- If the threshold is reached, go to **Step 3**.
- If the threshold is not reached, go to **Step 6**.

Step 3 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **Spark2x** > **Configurations** > **All Configurations** > **IndexServer2x** > **Tuning**. The default value of the **SPARK_DRIVER_MEMORY** is 4 GB. You can change the value according to the following rules: If this alarm is generated occasionally, increase the value by 0.5 times. Double the value if the alarm is reported frequently. In the case of large service volume and high service concurrency, you are advised to add instances.

Step 4 Restart all IndexServer2x instances.

Step 5 After 10 minutes, check whether the alarm is cleared.

- If the alarm is cleared, no further action is required.
- If the alarm is not cleared, go to **Step 6**.

Collect fault information.

Step 6 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Spark2x** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 9 Contact the O&M personnel and send the collected fault logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Reference

None

9.13.217 ALM-44004 Presto Coordinator Resource Group Queuing Tasks Exceed the Threshold

Description

This alarm is generated when the system detects that the number of queuing tasks in a resource group exceeds the threshold. The system queries the number of queuing tasks in a resource group through the JMX interface. You can choose **Components** > **Presto** > **Service Configuration** (switch **Basic** to **All**) > **Presto** > **resource-groups** to configure a resource group. You can choose **Components** > **Presto** > **Service Configuration** (switch **Basic** to **All**) > **Coordinator** > **Customize** > **resourceGroupAlarm** to configure the threshold of each resource group.

Attribute

Alarm ID	Alarm Severity	Auto Clear
44004	Major	Yes

Parameters

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

If the number of queuing tasks in a resource group exceeds the threshold, a large number of tasks may be in the queuing state. The Presto task time exceeds the expected value. When the number of queuing tasks in a resource group exceeds the maximum number (**maxQueued**) of queuing tasks in the resource group, new tasks cannot be executed.

Possible Causes

The resource group configuration is improper or too many tasks in the resource group are submitted.

Procedure

Step 1 Choose **Components** > **Presto** > **Service Configuration** (switch **Basic** to **All**) > **Presto** > **resource-groups** to adjust the resource group configuration.

Step 2 You can choose **Components** > **Presto** > **Service Configuration** (switch **Basic** to **All**) > **Coordinator** > **Customize** > **resourceGroupAlarm** to modify the threshold of each resource group.

Step 3 Collect the fault information.

1. Log in to the cluster node based on the host name in the fault information and query the number of queuing tasks based on **Resource Group** in the additional information on the Presto client.
2. Log in to the cluster node based on the host name in the fault information, view the **/var/log/Bigdata/nodeagent/monitorlog/monitor.log** file, and search for resource group information to view the monitoring collection information of the resource group.
3. Contact O&M personnel and send the collected logs.

----End

Related Information

None

9.13.218 ALM-44005 Presto Coordinator Process GC Time Exceeds the Threshold

Description

The system collects GC time of the Presto Coordinator process every 30 seconds. This alarm is generated when the GC time exceeds the threshold (exceeds 5 seconds for three consecutive times). You can change the threshold by choosing **System** > **Configure Alarm Threshold** > **Service** > **Presto** > **Coordinator** > **Presto Process Garbage Collection Time** > **Garbage Collection Time of the Coordinator Process** on MRS Manager. This alarm is cleared when the Coordinator process GC time is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
44005	Major	Yes

Parameter

Parameter	Description
ServiceName	Service for which the alarm is generated.

Parameter	Description
RoleName	Role for which the alarm is generated.
HostName	Host for which the alarm is generated.

Impact on the System

If the GC time of the Coordinator process is too long, the Coordinator process running performance will be affected and the Coordinator process will even be unavailable.

Possible Causes

The heap memory of the Coordinator process is overused or inappropriately allocated, causing frequent occurrence of the GC process.

Procedure

Step 1 Check the GC time.

1. Go to the cluster details page and choose **Alarms**.

   NOTE

   For MRS 1.8.10 or earlier, log in to MRS Manager and choose **Alarms**.

2. Select the alarm whose **Alarm ID** is **44005** and then check the role name in **Location** and confirm the IP adress of the instance.
3. Choose **Components** > **Presto** > **Instances** > **Coordinator** (business IP address of the instance for which the alarm is generated) > **Customize** > **Presto Garbage Collection Time**. Click **OK** to view the GC time.
4. Check whether the GC time of the Coordinator process is longer than 5 seconds.
   - If yes, go to **Step 1.5**.
   - If no, go to **Step 2**.
5. Choose **Components** > **Presto** > **Service Configuration**, and switch **Basic** to **All**. Choose **Presto** > **Coordinator**. Increase the value of **-Xmx** (maximum heap memory) in the **JAVA_OPTS** parameter based on the site requirements.
6. Check whether the alarm is cleared.
   - If yes, no further action is required.
   - If no, go to **Step 2**.

Step 2 Collect fault information.

1. On MRS Manager, choose **System** > **Export Log**.
2. Contact the O&M personnel and send the collected logs.

----End

Reference

None

9.13.219 ALM-44006 Presto Worker Process GC Time Exceeds the Threshold

Description

The system collects GC time of the Presto Worker process every 30 seconds. This alarm is generated when the GC time exceeds the threshold (exceeds 5 seconds for three consecutive times). You can change the threshold by choosing **System** > **Configure Alarm Threshold** > **Service** > **Presto** > **Worker** > **Presto Garbage Collection Time** > **Garbage Collection Time of the Worker Process** on MRS Manager. This alarm is cleared when the Worker process GC time is shorter than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
44006	Major	Yes

Parameter

Parameter	Description
ServiceName	Service for which the alarm is generated.
RoleName	Role for which the alarm is generated.
HostName	Host for which the alarm is generated.

Impact on the System

If the GC time of the Worker process is too long, the Worker process running performance will be affected and the Worker process will even be unavailable.

Possible Causes

The heap memory of the Worker process is overused or inappropriately allocated, causing frequent occurrence of the GC process.

Procedure

Step 1 Check the GC time.

1. Go to the cluster details page and choose **Alarms**.

NOTE

For MRS 1.8.10 or earlier, log in to MRS Manager and choose **Alarms**.

2. Select the alarm whose **Alarm ID** is **44006**. Then check the role name in **Location** and confirm the IP adress of the instance.
3. Choose **Components** > **Presto** > **Instances** > **Worker** (business IP address of the instance for which the alarm is generated) > **Customize** > **Presto Garbage Collection Time**. Click **OK** to view the GC time.
4. Check whether the GC time of the Worker process is longer than 5 seconds.
   - If yes, go to **Step 1.5**.
   - If no, go to **Step 2**.
5. Choose **Components** > **Presto** > **Service Configuration**, and switch **Basic** to **All**, and choose **Presto** > **Worker** Increase the value of **-Xmx** (maximum heap memory) in the **JAVA_OPTS** parameter based on the site requirements.
6. Check whether the alarm is cleared.
   - If yes, no further action is required.
   - If no, go to **Step 2**.

Step 2 Collect fault information.

1. On MRS Manager, choose **System** > **Export Log**.
2. Contact the O&M personnel and send the collected logs.

----End

Reference

None

9.13.220 ALM-45000 HetuEngine Service Unavailable

Description

The system checks the HetuEngine service status every 300 seconds. This alarm is generated when the HetuEngine service is unavailable.

This alarm is cleared when the HetuEngine service recovers.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45000	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

HetuEngine tasks fail to execute.

Possible Causes

- The KrbServer service is abnormal.
- The ZooKeeper service is abnormal.
- The HDFS service is abnormal.
- The Yarn service is abnormal.
- The DBService service is abnormal.
- The Hive service is abnormal.
- Thre are no HSBroker instances in HetuEngine.

Procedure

Check the KrbServer service status.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarm**.

Step 2 In the alarm list, check whether the "ALM-25500 KrbServer Service Unavailable" alarm is generated.

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 Clear "ALM-25500 KrbServer Service Unavailable" according to the alarm help.

Step 4 In the alarm list, check whether the alarm "ALM-45000 HetuEngine Service Unavailable" is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Check the ZooKeeper service status.

Step 5 In the alarm list, check whether the alarm "ALM-12007 Process Fault" is generated.

- If yes, go to **Step 6**.
- If no, go to **Step 9**.

Step 6 In the alarm list, click ⌄ in the row that contains the "Process Fault" alarm. Check whether the name of the service for which the alarm is generated is ZooKeeper in **Location Information**.

- If yes, go to **Step 7**.
- If no, go to **Step 9**.

Step 7 Clear "ALM-12007 Process Fault" according to the alarm help.

Step 8 In the alarm list, check whether the alarm "ALM-45000 HetuEngine Service Unavailable" is cleared.

- If yes, no further action is required.
- If no, go to **Step 9**.

Check the HDFS service status.

Step 9 In the alarm list, check whether the "ALM-14000 HDFS Service Unavailable" alarm is generated.

- If yes, go to **Step 10**.
- If no, go to **Step 12**.

Step 10 Clear "ALM-14000 HDFS Service Unavailable" according to the alarm help.

Step 11 In the alarm list, check whether the "ALM-45000 HetuEngine Service Unavailable" alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 12**.

Check the YARN service status.

Step 12 In the alarm list, check whether the "ALM-18000 YARN Service Unavailable" alarm is generated.

- If yes, go to **Step 13**.
- If no, go to **Step 15**.

Step 13 Clear "ALM-18000 YARN Service Unavailable" according to the alarm help.

Step 14 In the alarm list, check whether the "ALM-45000 HetuEngine Service Unavailable" alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 15**.

Check the DBService service status.

Step 15 In the alarm list, check whether the "ALM-27001 DBService Service Unavailable" alarm is generated.

- If yes, go to **Step 16**.
- If no, go to **20**.

Step 16 Clear "ALM-27001 DBService Service Unavailable" according to the alarm help.

Step 17 In the alarm list, check whether the "ALM-45000 HetuEngine Service Unavailable" alarm is cleared.

- If yes, no further action is required.
- If no, go to 20.

Check the Hive service status.

Step 18 In the alarm list, check whether the "ALM-16004 Hive Service Unavailable" alarm is generated.

- If yes, go to **Step 19**.
- If no, go to 20.

Step 19 Clear "ALM-16004 Hive Service Unavailable" according to the alarm help.

Step 20 In the alarm list, check whether the "ALM-45000 HetuEngine Service Unavailable" alarm is cleared.

- If yes, no further action is required.
- If no, go to 20.

Check whether there are no HSBroker instances in HetuEngine.

Step 21 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **HetuEngine**. On the page that is displayed, click the **Instance** tab.

Step 22 Check whether there are no HSBroker instances.

- If yes, click **Add Instance** to add one.
- If no, go to 23.

Step 23 In the alarm list, check whether the "ALM-45000 HetuEngine Service Unavailable" alarm is cleared.

- If yes, no further action is required.
- If no, go to 23.

Check the network connection between HetuEngine and ZooKeeper, HDFS, YARN, DBService, and Hive.

Step 24 On FusionInsight Manager, choose **Cluster** > *Name of the desired cluster* > **Services** > **HetuEngine**. On the page that is displayed, click the **Instance** tab.

Step 25 Click the host name in the **HSBroker** row and record the management IP address in the **Basic Information** area.

Step 26 Log in to the host where HSBroker resides as user **omm** using the IP address obtained in **Step 25**.

Step 27 Run the **ping** command to check whether the network connection between the host where HSBroker resides and the hosts where ZooKeeper, HDFS, Yarn, DBService, and Hive reside is in the normal state.

- If yes, go to **Step 30**.
- If no, go to **Step 28**.

Step 28 Contact the network administrator to restore the network.

Step 29 In the alarm list, check whether the "ALM-45000 HetuEngine Service Unavailable" alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 30**.

Collect fault information.

Step 30 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 31 Expand the **Service** drop-down list. In the **Services** dialog box that is displayed, select **HetuEngine** under the target cluster name, and click **OK**.

Step 32 Expand the **Hosts** drop-down list. In the **Select Host** dialog box that is displayed, select the hosts to which the role belongs, and click **OK**.

Step 33 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 34 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

After the fault is rectified, the system automatically clears this alarm.

Reference

None

9.13.221 ALM-45175 Average Time for Calling OBS Metadata APIs Is Greater than the Threshold

Description

The system checks whether the average duration for calling OBS metadata APIs is greater than the threshold every 30 seconds. This alarm is generated when the number of consecutive times that the average time exceeds the specified threshold is greater than the number of smoothing times.

This alarm is automatically cleared when the average duration for calling the OBS metadata APIs is lower than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45175	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

If the average time for calling the OBS metadata APIs exceeds the threshold, the upper-layer big data computing services may be affected. To be more specific, the execution time of some computing tasks will exceed the threshold.

Possible Causes

Frame freezing occurs on the OBS server, or the network between the OBS client and the OBS server is unstable.

Procedure

Check the heap memory usage.

Step 1 On the **FusionInsight Manager** homepage, choose **O&M** > **Alarm** > **Alarms** > **Average Time for Calling the OBS Metadata API Exceeds the Threshold**, view the role name in **Location**, and check the instance IP address.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services** > **meta** > **Instance** > **meta** (IP address of the instance for which the alarm is generated). Click the drop-down list in the upper right corner of the chart area and choose **Customize**. In the dialog box that is displayed, select **Average time of OBS interface calls** from **OBS Meta data Operations**, and click **OK**. Check whether the average time of OBS metadata API calls exceeds the threshold.

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 Choose **Cluster** > *Name of the desired cluster* > **O&M** > **Alarm** > **Thresholds** > **meta** > **Average Time for Calling the OBS Metadata API**. Increase the threshold or smoothing times as required.

Step 4 Check whether the alarm is cleared.

- If yes, no further action is required.

- If no, go to **Step 5**.

Collect the fault information.

Step 5 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 6 In the **Services** area, select **NodeAgent**, **NodeMetricAgent**, **OmmServer**, and **OmmAgent** under OMS.

Step 7 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 8 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.222 ALM-45176 Success Rate of Calling OBS Metadata APIs Is Lower than the Threshold

Description

The system checks whether the success rate of calling OBS metadata APIs is lower than the threshold every 30 seconds. This alarm is generated when the success rate is lower than the threshold.

This alarm is automatically cleared when the success rate of calling APIs for writing OBS data is greater than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45176	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

If the success rate of calling the OBS metadata APIs is less than the threshold, the upper-layer big data computing services may be affected. To be more specific, some computing tasks may fail to be executed.

Possible Causes

An execution exception or severe timeout occurs on the OBS server.

Procedure

Check the heap memory usage.

Step 1 On the **FusionInsight Manager** homepage, choose **O&M** > **Alarm** > **Alarms** > **Success Rate for Calling the OBS Metadata API Is Lower Than the Threshold**, view the role name in **Location**, and check the instance IP address.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services** > **meta** > **Instance** > **meta** (IP address of the instance for which the alarm is generated). Click the drop-down list in the upper right corner of the chart area and choose **Customize**. In the dialog box that is displayed, select **Success percent of OBS interface calls** from **OBS Meta data Operations**, and click **OK**. Check whether the average time of OBS metadata API calls exceeds the threshold.

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 Choose **Cluster** > *Name of the desired cluster* > **O&M** > **Alarm** > **Thresholds** > **meta** > **Success Rate for Calling the OBS Metadata API**. Increase the threshold or smoothing times as required.

Step 4 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Collect the fault information.

Step 5 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 6 In the **Services** area, select **NodeAgent**, **NodeMetricAgent**, **OmmServer**, and **OmmAgent** under OMS.

Step 7 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 8 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.223 ALM-45177 Success Rate of Calling OBS Data Read APIs Is Lower than the Threshold

Description

The system checks whether the success rate of calling APIs for reading OBS data is lower than the threshold every 30 seconds. This alarm is generated when the success rate is lower than the threshold.

This alarm is automatically cleared when the success rate of calling APIs for reading OBS data is greater than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45177	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Name	Meaning
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

If the success rate of calling the OBS APIs for reading data is less than the threshold, the upper-layer big data computing services may be affected. To be more specific, some computing tasks may fail to be executed.

Possible Causes

An execution exception or severe timeout occurs on the OBS server.

Procedure

Check the heap memory usage.

Step 1 On the **FusionInsight Manager** homepage, choose **O&M** > **Alarm** > **Alarms** > **Success Rate for Calling the OBS Data Read API Is Lower Than the Threshold**, view the role name in **Location**, and check the instance IP address.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services** > **meta** > **Instance** > **meta** (IP address of the instance for which the alarm is generated). Click the drop-down list in the upper right corner of the chart area and choose **Customize**. In the dialog box that is displayed, select **Success percent of OBS data read operation interface calls** from **OBS data read operation**, and click **OK**. Check whether the average time of OBS metadata API calls exceeds the threshold.

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 Choose **Cluster** > *Name of the desired cluster* > **O&M** > **Alarm** > **Thresholds** > **meta** > **Success Rate for Calling the OBS Data Read API**. Increase the threshold or smoothing times as required.

Step 4 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Collect the fault information.

Step 5 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 6 In the **Services** area, select **NodeAgent**, **NodeMetricAgent**, **OmmServer**, and **OmmAgent** under OMS.

Step 7 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 8 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.224 ALM-45178 Success Rate of Calling OBS Data Write APIs Is Lower Than the Threshold

Description

The system checks whether the success rate of calling APIs for writing OBS data is lower than the threshold every 30 seconds. This alarm is generated when the success rate is lower than the threshold.

This alarm is automatically cleared when the success rate of calling APIs for writing OBS data is greater than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45178	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

If the success rate of calling the OBS APIs for writing data is lower than the threshold, the upper-layer big data computing services may be affected. To be more specific, some computing tasks may fail to be executed.

Possible Causes

An execution exception or severe timeout occurs on the OBS server.

Procedure

Check the heap memory usage.

Step 1 On the **FusionInsight Manager** homepage, choose **O&M** > **Alarm** > **Alarms** > **Success Rate for Calling the OBS Data Write API Is Lower Than the Threshold**, view the role name in **Location**, and check the instance IP address.

Step 2 Choose **Cluster** > *Name of the desired cluster* > **Services** > **meta** > **Instance** > **meta** (IP address of the instance for which the alarm is generated). Click the drop-down list in the upper right corner of the chart area and choose **Customize**. In the dialog box that is displayed, select **Success percent of OBS data write operation interface calls** from **OBS data write operation**, and click **OK**. Check whether the average time of OBS metadata API calls exceeds the threshold.

- If yes, go to **Step 3**.
- If no, go to **Step 5**.

Step 3 Choose **Cluster** > *Name of the desired cluster* > **O&M** > **Alarm** > **Thresholds** > **meta** > **Success Rate for Calling the OBS Data Write API**. Increase the threshold or smoothing times as required.

Step 4 Check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Collect the fault information.

Step 5 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 6 In the **Services** area, select **NodeAgent**, **NodeMetricAgent**, **OmmServer**, and **OmmAgent** under OMS.

Step 7 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 8 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.225 ALM-45275 Ranger Service Unavailable

Description

The alarm module checks the Ranger service status every 180 seconds. This alarm is generated if the Ranger service is abnormal.

This alarm is cleared after the Ranger service recovers.

Attributes

Alarm ID	Alarm Severity	Automatically Cleared
45275	Critical	Yes

Parameters

Name	Meaning
Source	Cluster for which the alarm is generated.
ServiceName	Service for which the alarm is generated.
RoleName	Role for which the alarm is generated.
HostName	Host for which the alarm is generated.

Impact on the System

When the Ranger service is unavailable, Ranger cannot work properly and the native Ranger UI cannot be accessed.

Possible Causes

- The DBService service on which Ranger depends is abnormal.
- The RangerAdmin role instance is abnormal.

Procedure

Check the DBService process status.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. On the displayed page, check whether the ALM-27001 DBService Service Unavailable alarm is reported.

- If yes, go to **Step 2**.
- If no, go to **Step 3**.

Step 2 Rectify the DBService service fault by following the handling procedure of ALM-27001 DBService Service Unavailable. After the DBService alarm is cleared, check whether Ranger Service Unavailable alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 3**.

Check all RangerAdmin instances.

Step 3 Log in to the node where the RangerAdmin instance is located as user **omm** and run the **ps -ef|grep "proc_rangeradmin"** command to check whether the RangerAdmin process exists on the current node.

- If yes, go to **Step 5**.
- If no, restart the faulty RangerAdmin instance or Ranger service and go to **Step 4**.

Step 4 In the alarm list, check whether the alarm "Ranger Service Unavailable" is cleared.

- If yes, no further action is required.
- If no, go to **Step 5**.

Collect the fault information.

Step 5 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 6 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 7 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 8 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

After the fault that triggers the alarm is rectified, the alarm is automatically cleared.

Related Information

None

9.13.226 ALM-45276 Abnormal RangerAdmin Status

Description

The alarm module checks the RangerAdmin service status every 60 seconds. This alarm is generated if RangerAdmin is unavailable.

This alarm is automatically cleared after the RangerAdmin service recovers.

Attributes

Alarm ID	Alarm Severity	Automatically Cleared
45276	Major	Yes

Parameters

Name	Meaning
Source	Cluster for which the alarm is generated.
ServiceName	Service for which the alarm is generated.
RoleName	Role for which the alarm is generated.
HostName	Host for which the alarm is generated.

Impact on the System

If the status of a RangerAdmin is abnormal, access to the Ranger native UI is not affected. If there are two abnormal RangerAdmin instances, the Ranger native UI cannot be accessed and operations such as creating, modifying, and deleting policies are unavailable.

Possible Causes

The RangerAdmin port is not started.

Procedure

Check the port process.

Step 1 In the alarm list on FusionInsight Manager, locate the row that contains the alarm, and click ⌄ to view the name of the host for which the alarm is generated.

Step 2 Log in to the node where the RangerAdmin instance is located as user **omm**. Run the **ps -ef|grep "proc_rangeradmin" | grep -v grep | awk -F ' ' '{print $2}'** command to obtain ***pid*** of the RangerAdmin process, and run the **netstat -anp| grep *pid* | grep LISTEN** command to check whether the RangerAdmin process listens to port 21401 in the security mode and port 21400 in standard mode.

- If yes, go to **Step 4**.
- If no, restart the faulty RangerAdmin instance or Ranger service and go to **Step 3**.

Step 3 In the alarm list, check whether the "Abnormal RangerAdmin Status" alarm is cleared.

- If yes, no further action is required.

- If no, go to **Step 4**.

Collect the fault information.

Step 4 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 5 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 6 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 7 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

After the fault that triggers the alarm is rectified, the alarm is automatically cleared.

Related Information

None

9.13.227 ALM-45277 RangerAdmin Heap Memory Usage Exceeds the Threshold

Description

The system checks the heap memory usage of the RangerAdmin service every 60 seconds. This alarm is generated when the system detects that the heap memory usage of the RangerAdmin instance exceeds the threshold (95% of the maximum memory) for 10 consecutive times. This alarm is cleared when the heap memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45277	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.

Name	Meaning
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Heap memory overflow may cause service breakdown.

Possible Causes

The heap memory usage of the RangerAdmin instance is high or the heap memory is improperly allocated.

Procedure

Check the heap memory usage.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **ALM-45277 RangerAdmin Heap Memory Usage Exceeds the Threshold**. Check the location information of the alarm and view the host name of the instance for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance**. Select the role corresponding to the host name of the instance for which the alarm is generated. Click the drop-down list in the upper right corner of the chart area and choose **Customize** > **CPU and Memory** > **RangerAdmin Heap Memory Usage**. Click **OK**.

Step 3 Check whether the heap memory used by RangerAdmin reaches the threshold (95% of the maximum heap memory by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance** > **RangerAdmin** > **Instance Configuration**. Click **All Configurations**, and choose **RangerAdmin** > **System**. Increase the value of **-Xmx** in the **GC_OPTS** parameter based on the site requirements and save the configuration.

NOTE

If this alarm is generated, the heap memory configured for RangerAdmin cannot meet the heap memory required by the RangerAdmin process. You are advised to check the heap memory usage of RangerAdmin and change the value of **-Xmx** in **GC_OPTS** to the twice of the heap memory used by RangerAdmin. The value can be changed based on the actual service scenario. For details, see **Step 2**.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.228 ALM-45278 RangerAdmin Direct Memory Usage Exceeds the Threshold

Description

The system checks the direct memory usage of the RangerAdmin service every 60 seconds. This alarm is generated when the direct memory usage of the RangerAdmin instance exceeds the threshold (80% of the maximum memory) for five consecutive times. This alarm is cleared when the direct memory usage of RangerAdmin is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45278	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.

Name	Meaning
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Direct memory overflow may cause service breakdown.

Possible Causes

The direct memory of the RangerAdmin instance is overused or the direct memory is inappropriately allocated. As a result, the memory usage exceeds the threshold.

Procedure

Check the direct memory usage.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **ALM-45278 RangerAdmin Direct Memory Usage Exceeds the Threshold**. Check the location information of the alarm and view the host name of the instance for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance**. Select the role corresponding to the host name of the instance for which the alarm is generated. Click the drop-down list in the upper right corner of the chart area and choose **Customize** > **CPU and Memory** > **RangerAdmin Direct Memory Usage**. Click **OK**.

Step 3 Check whether the direct memory used by RangerAdmin reaches the threshold (80% of the maximum direct memory by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance** > **RangerAdmin** > **Instance Configuration**. Click **All Configurations**, and choose **RangerAdmin** > **System**. Increase the value of **-XX:MaxDirectMemorySize** in the **GC_OPTS** parameter based on the site requirements and save the configuration.

NOTE

If this alarm is generated, the direct memory configured for RangerAdmin cannot meet the direct memory required by the RangerAdmin process. You are advised to check the direct memory usage of RangerAdmin and change the value of **-XX:MaxDirectMemorySize** in **GC_OPTS** to the twice of the direct memory used by RangerAdmin. You can change the value based on the actual service scenario. For details, see **Step 2**.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.229 ALM-45279 RangerAdmin Non Heap Memory Usage Exceeds the Threshold

Description

The system checks the non-heap memory usage of the RangerAdmin service every 60 seconds. This alarm is generated when the non-heap memory usage of the RangerAdmin instance exceeds the threshold (80% of the maximum memory) for five consecutive times. This alarm is cleared when the non-heap memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45279	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Non-heap memory overflow may cause service breakdown.

Possible Causes

The non-heap memory usage of the RangerAdmin instance is high or the non-heap memory is improperly allocated.

Procedure

Check non-heap memory usage.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **ALM-45279 RangerAdmin Non Heap Memory Usage Exceeds the Threshold**. Check the location information of the alarm and view the host name of the instance for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance**. Select the role corresponding to the host name of the instance for which the alarm is generated. Click the drop-down list in the upper right corner of the chart area and choose **Customize** > **CPU and Memory** > **RangerAdmin Non Heap Memory Usage**. Click **OK**.

Step 3 Check whether the non-heap memory used by RangerAdmin reaches the threshold (80% of the maximum non-heap memory by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance** > **RangerAdmin** > **Instance Configuration**. Click **All Configurations**, and choose **RangerAdmin** > **System**. Set **-XX: MaxPermSize** in the **GC_OPTS** parameter to a larger value based on site requirements and save the configuration.

NOTE

If this alarm is generated, the non-heap memory size configured for the RangerAdmin instance cannot meet the non-heap memory required by the RangerAdmin process. You are advised to change the value of **-XX:MaxPermSize** in **GC_OPTS** to the twice of the current non-heap memory usage or change the value based on the site requirements.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.230 ALM-45280 RangerAdmin GC Duration Exceeds the Threshold

Description

The system checks the GC duration of the RangerAdmin process every 60 seconds. This alarm is generated when the GC duration of the RangerAdmin process exceeds the threshold (12 seconds by default) for five consecutive times. This alarm is cleared when the GC duration is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45280	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

The RangerAdmin responds slowly.

Possible Causes

The heap memory of the RangerAdmin instance is overused or the heap memory is inappropriately allocated. As a result, GCs occur frequently.

Procedure

Check the GC duration.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **ALM-45280 RangerAdmin GC Duration Exceeds the Threshold**. Check the location information of the alarm and view the host name of the instance for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance**. Select the role corresponding to the host name of the instance for which the alarm is generated and click the drop-down list in the upper right corner of the chart area. Choose **Customize** > **GC** > **RangerAdmin GC Duration**. Click **OK**.

Step 3 Check whether the GC duration of the RangerAdmin process collected every minute exceeds the threshold (12 seconds by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance** > **RangerAdmin** > **Instance Configuration**. Click **All Configurations**, and choose **RangerAdmin** > **System**. Increase the value of **-Xmx** in the **GC_OPTS** parameter based on the site requirements and save the configuration.

NOTE

> If this alarm is generated, the heap memory configured for RangerAdmin cannot meet the heap memory required by the RangerAdmin process. You are advised to check the heap memory usage of RangerAdmin and change the value of **-Xmx** in **GC_OPTS** to the twice of the heap memory used by RangerAdmin. The value can be changed based on the actual service scenario. For details, see **Step 2**.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.231 ALM-45281 UserSync Heap Memory Usage Exceeds the Threshold

Description

The system checks the heap memory usage of the UserSync service every 60 seconds. This alarm is generated when the system detects that the heap memory usage of the UserSync instance exceeds the threshold (95% of the maximum memory) for 10 consecutive times. This alarm is cleared when the heap memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45281	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Heap memory overflow may cause service breakdown.

Possible Causes

The heap memory usage of the UserSync instance is high or the heap memory is improperly allocated.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **ALM-45281 UserSync Heap Memory Usage Exceeds the Threshold**. Check the location information of the alarm and view the host name of the instance for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance**. Select the role corresponding to the host name of the instance for which the alarm is generated. Click the drop-down list in the upper right corner of the chart area and choose **Customize** > **CPU and Memory** > **UserSync Heap Memory Usage**. Click **OK**.

Step 3 Check whether the heap memory used by UserSync reaches the threshold (95% of the maximum heap memory by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance** > **UserSync** > **Instance Configuration**. Click **All Configurations**, and choose **UserSync** > **System**. Increase the value of **-Xmx** in the **GC_OPTS** parameter based on the site requirements and save the configuration.

NOTE

If this alarm is generated, the heap memory configured for UserSync cannot meet the heap memory required by the UserSync process. You are advised to change the **-Xmx** value of **GC_OPTS** to twice that of the heap memory used by UserSync. You can change the value based on the actual service scenario. For details about how to check the UserSync heap memory usage, see **Step 2**.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.232 ALM-45282 UserSync Direct Memory Usage Exceeds the Threshold

Description

The system checks the direct memory usage of the UserSync service every 60 seconds. This alarm is generated when the direct memory usage of the UserSync instance exceeds the threshold (80% of the maximum memory) for five consecutive times. This alarm is cleared when the UserSync direct memory usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45282	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Direct memory overflow may cause service breakdown.

Possible Causes

The direct memory of the UserSync instance is overused or the direct memory is inappropriately allocated. As a result, the memory usage exceeds the threshold.

Procedure

Check the direct memory usage.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **ALM-45282 UserSync Direct Memory Usage Exceeds the Threshold**. Check the location information of the alarm. Check the name of the instance host for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance**. Select the role corresponding to the host name of the instance for which the alarm is generated. Click the drop-down list in the upper right corner of the chart area and choose **Customize** > **CPU and Memory** > **UserSync Direct Memory Usage**. Click **OK**.

Step 3 Check whether the direct memory used by the UserSync reaches the threshold (80% of the maximum direct memory by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance** > **UserSync** > **Instance Configuration**. Click **All Configurations**, and choose **UserSync** > **System**. Increase the value of **-XX:MaxDirectMemorySize** in the **GC_OPTS** parameter based on the site requirements and save the configuration.

NOTE

If this alarm is generated, the direct memory configured for UserSync cannot meet the direct memory required by the UserSync process. You are advised to check the direct memory usage of UserSync and change the value of **-XX:MaxDirectMemorySize** in **GC_OPTS** to the twice of the direct memory used by UserSync. You can change the value based on the actual service scenario. For details, see **Step 2**.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.233 ALM-45283 UserSync Non Heap Memory Usage Exceeds the Threshold

Description

The system checks the non-heap memory usage of the UserSync service every 60 seconds. This alarm is generated when the non-heap memory usage of the UserSync instance exceeds the threshold (80% of the maximum memory) for five consecutive times. This alarm is cleared when the non-heap memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45283	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Non-heap memory overflow may cause service breakdown.

Possible Causes

The non-heap memory of the UserSync process is overused or the non-heap memory is inappropriately allocated.

Procedure

Check non-heap memory usage.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **ALM-45283 UserSync Non Heap Memory Usage Exceeds the Threshold**. Check the location information of the alarm and view the host name of the instance for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance**. Select the role corresponding to the host name of the instance for which the alarm is generated. Click the drop-down list in the upper right corner of the chart area and choose **Customize** > **CPU and Memory** > **UserSync Non Heap Memory Usage**. Click **OK**.

Step 3 Check whether the non-heap memory used by UserSync reaches the threshold (80% of the maximum non-heap memory by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance** > **UserSync** > **Instance Configuration**. Click **All Configurations**, and choose **UserSync** > **System**. Set **-XX: MaxPermSize** in the **GC_OPTS** parameter to a larger value based on site requirements and click **Save** to save the configuration.

NOTE

If this alarm is generated, the non-heap memory size configured for the UserSync instance cannot meet the non-heap memory required by the UserSync process. You are advised to change the **-XX:MaxPermSize** value of **GC_OPTS** to twice that of the current non-heap memory size or change the value based on the site requirements.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.234 ALM-45284 UserSync Garbage Collection (GC) Time Exceeds the Threshold

Description

The system checks the GC duration of the UserSync process every 60 seconds. This alarm is generated when the GC duration of the UserSync process exceeds the threshold (12 seconds by default) for five consecutive times. This alarm is cleared when the GC duration is less than the threshold.

Attributes

Alarm ID	Alarm Severity	Automatically Cleared
45284	Major	Yes

Parameters

Name	Meaning
Source	Cluster for which the alarm is generated.
ServiceName	Service for which the alarm is generated.
RoleName	Role for which the alarm is generated.
HostName	Host for which the alarm is generated.
Trigger Condition	Threshold for triggering the alarm.

Impact on the System

UserSync responds slowly.

Possible Causes

The heap memory of the UserSync instance is overused or the heap memory is inappropriately allocated. As a result, GCs occur frequently.

Procedure

Check the GC time.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **ALM-45284 UserSync Garbage Collection (GC) Time Exceeds the Threshold**. Check the location information of the alarm and view the host name of the instance for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance**. Select the role corresponding to the host name of the instance for which the alarm is generated and click the drop-down list in the upper right corner of the chart area. Choose **Customize** > **GC** > **UserSync GC Duration**. Click **OK**.

Step 3 Check whether the GC duration of the UserSync process collected every minute exceeds the threshold (12 seconds by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance** > **UserSync** > **Instance Configuration**. Click **All Configurations**, and choose **UserSync** > **System**. Increase the value of **-Xmx** in the **GC_OPTS** parameter based on the site requirements and save the configuration.

NOTE

> If this alarm is generated, the heap memory configured for UserSync cannot meet the heap memory required by the UserSync process. You are advised to change the value of **GC_OPTS** to the twice that of the heap memory used by UserSync. You can change the value based on the actual service scenario. For details about how to check the UserSync heap memory usage, see **Step 2**.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M** > **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

After the fault that triggers the alarm is rectified, the alarm is automatically cleared.

Related Information

None

9.13.235 ALM-45285 TagSync Heap Memory Usage Exceeds the Threshold

Description

The system checks the heap memory usage of the TagSync service every 60 seconds. This alarm is generated when the heap memory usage of the TagSync instance exceeds the threshold (95% of the maximum memory) for 10 consecutive times. This alarm is cleared when the heap memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45285	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Heap memory overflow may cause service breakdown.

Possible Causes

The heap memory usage of the TagSync instance is high or the heap memory is improperly allocated.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **ALM-45285 TagSync Heap Memory Usage Exceeds the Threshold**. Check the location information of the alarm and view the host name of the instance for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance**. Select the role corresponding to the host name of the instance for which the alarm is generated. Click the drop-down list in the upper right corner of the chart area and choose **Customize** > **CPU and Memory** > **TagSync Heap Memory Usage**. Click **OK**.

Step 3 Check whether the heap memory used by TagSync reaches the threshold (95% of the maximum heap memory by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance** > **TagSync** > **Instance Configuration**. Click **All Configurations** and choose **TagSync** > **System**. Increase the value of **-Xmx** in the **GC_OPTS** parameter based on the site requirements and save the configuration.

> **NOTE**
>
> If this alarm is generated, the heap memory configured for TagSync cannot meet the heap memory required by the TagSync process. You are advised to change the **-Xmx** value of **GC_OPTS** to twice that of the heap memory used by TagSync. You can change the value based on the actual service scenario. For details about how to check the TagSync heap memory usage, see **Step 2**.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.236 ALM-45286 TagSync Direct Memory Usage Exceeds the Threshold

Description

The system checks the direct memory usage of the TagSync service every 60 seconds. This alarm is generated when the direct memory usage of the TagSync instance exceeds the threshold (80% of the maximum memory) for five consecutive times. This alarm is cleared when the TagSync direct memory usage is less than or equal to the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45286	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Direct memory overflow may cause service breakdown.

Possible Causes

The direct memory of the TagSync instance is overused or the direct memory is inappropriately allocated. As a result, the memory usage exceeds the threshold.

Procedure

Check the direct memory usage.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **ALM-45286 TagSync Direct Memory Usage Exceeds the Threshold**. Check the location information of the alarm and view the host name of the instance for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance**. Select the role corresponding to the host name of the instance for which the alarm is generated. Click the drop-down list in the upper right corner of the chart area and choose **Customize** > **CPU and Memory** > **TagSync Direct Memory Usage**. Click **OK**.

Step 3 Check whether the direct memory used by the TagSync reaches the threshold (80% of the maximum direct memory by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance** > **TagSync** > **Instance Configuration**. Click **All Configurations** and choose **TagSync** > **System**. Increase the value of **-XX:MaxDirectMemorySize** in the **GC_OPTS** parameter based on the site requirements and save the configuration.

NOTE

If this alarm is generated, the direct memory configured for TagSync cannot meet the direct memory required by the TagSync process. You are advised to check the direct memory usage of TagSync and change the value of **-XX:MaxDirectMemorySize** in **GC_OPTS** to the twice of the direct memory used by TagSync. You can change the value based on the actual service scenario. For details, see **Step 2**.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.237 ALM-45287 TagSync Non Heap Memory Usage Exceeds the Threshold

Description

The system checks the non-heap memory usage of the TagSync service every 60 seconds. This alarm is generated when the non-heap memory usage of the TagSync instance exceeds the threshold (80% of the maximum memory) for five consecutive times. This alarm is cleared when the non-heap memory usage is less than the threshold.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45287	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

Non-heap memory overflow may cause service breakdown.

Possible Causes

The non-heap memory of the TagSync process is overused or the non-heap memory is inappropriately allocated.

Procedure

Check non-heap memory usage.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **ALM-45287 TagSync Non Heap Memory Usage Exceeds the Threshold**. Check the location information of the alarm and view the host name of the instance for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance**. Select the role corresponding to the host name of the instance for which the alarm is generated. Click the drop-down list in the upper right corner of the chart area and choose **Customize** > **CPU and Memory** > **TagSync Non Heap Memory Usage**. Click **OK**.

Step 3 Check whether the non-heap memory used by TagSync reaches the threshold (80% of the maximum non-heap memory by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance** > **TagSync** > **Instance Configuration**. Click **All Configurations** and choose **TagSync** > **System**. Set **-XX: MaxPermSize** in the **GC_OPTS** parameter to a larger value based on site requirements and save the configuration.

NOTE

If this alarm is generated, the non-heap memory size configured for the TagSync instance cannot meet the non-heap memory required by the TagSync process. You are advised to change the **-XX:MaxPermSize** value of **GC_OPTS** to twice that of the current non-heap memory size or change the value based on the site requirements.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.238 ALM-45288 TagSync Garbage Collection (GC) Time Exceeds the Threshold

Description

The system checks the GC duration of the TagSync process every 60 seconds. This alarm is generated when the GC duration of the TagSync process exceeds the threshold (12 seconds by default) for five consecutive times. This alarm is cleared when the GC duration is less than the threshold.

Attributes

Alarm ID	Alarm Severity	Auto Clear
45288	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.
Trigger Condition	Specifies the threshold for triggering the alarm.

Impact on the System

TagSync responds slowly.

Possible Causes

The heap memory of the TagSync instance is overused or the heap memory is inappropriately allocated. As a result, GCs occur frequently.

Procedure

Check the GC duration.

Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **ALM-45288 TagSync Garbage Collection (GC) Time Exceeds the Threshold**. Check the location information of the alarm and view the host name of the instance for which the alarm is generated.

Step 2 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance**. Select the role corresponding to the host name of the instance for which the alarm is generated and click the drop-down list in the upper right corner of the chart area. Choose **Customize** > **GC** > **TagSync GC Duration**. Click **OK**.

Step 3 Check whether the GC duration of the TagSync process collected every minute exceeds the threshold (12 seconds by default).

- If yes, go to **Step 4**.
- If no, go to **Step 6**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **Ranger** > **Instance** > **TagSync** > **Instance Configuration**. Click **All Configurations** and choose **TagSync** > **System**. Increase the value of **-Xmx** in the **GC_OPTS** parameter based on the site requirements and save the configuration.

NOTE

> If this alarm is generated, the heap memory configured for TagSync cannot meet the heap memory required by the TagSync process. You are advised to change the **-Xmx** value of **GC_OPTS** to twice that of the heap memory used by TagSync. You can change the value based on the actual service scenario. For details about how to check the TagSync heap memory usage, see **Step 2**.

Step 5 Restart the affected services or instances and check whether the alarm is cleared.

- If yes, no further action is required.
- If no, go to **Step 6**.

Collect the fault information.

Step 6 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 7 Expand the **Service** drop-down list, and select **Ranger** for the target cluster.

Step 8 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 10 minutes ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 9 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.239 ALM-45425 ClickHouse Service Unavailable

Description

The alarm module checks the ClickHouse instance status every 60 seconds. This alarm is generated when the alarm module detects that all ClickHouse instances are abnormal.

This alarm is cleared when the system detects that any ClickHouse instance is restored and the alarm is cleared.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45425	Critical	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The ClickHouse service is abnormal. You cannot use FusionInsight Manager to perform cluster operations on the ClickHouse service. The ClickHouse service function is unavailable.

Possible Causes

The configuration information in the **metrika.xml** file in the component configuration directory of the faulty ClickHouse instance node is inconsistent with that of the corresponding ClickHouse instance in the ZooKeeper.

Procedure

Check whether the configuration in metrika.xml of the ClickHouse instance is correct.

Step 1 Log in to FusionInsight Manager, choose **Cluster** > **Services** > **ClickHouse** > **Instance**, and locate the abnormal ClickHouse instance based on the alarm information.

- If yes, go to **Step 2**.
- If no, go to **Step 9**.

Step 2 Log in to the host where the ClickHouse service is abnormal and ping the IP address of another normal ClickHouse instance node to check whether the network connection is normal.

- If yes, go to **Step 3**.
- If no, contact the network administrator to repair the network.

Step 3 Choose **Cluster** > **Services** > **ClickHouse** > **Instance**, click the abnormal instance name in the **Role** column, click **Configurations**, search for **macros.id** in the search box, and find the value of **macros.id** of the current instance.

Step 4 Log in to the host where the ZooKeeper client is located and log in to the ZooKeeper client.

Switch to the client installation directory.

Example: **cd /opt/client**

Run the following command to configure environment variables:

source bigdata_env

Run the following command to authenticate the user (skip this step in common mode):

kinit *Component service user*

Run the following command to log in to the client tool:

zkCli.sh -server *service IP address of the node where the ZooKeeper role instance locates*:*client port*

Step 5 Run the following command to check whether the ClickHouse cluster topology information can be obtained.

get /clickhouse/config/*value of* ***macros.id*** *in* **Step 3/metrika.xml**

- If yes, go to **Step 6**.
- If no, go to **Step 9**.

Step 6 Log in to the host where the ClickHouse instance is abnormal and go to the configuration directory of the ClickHouse instance.

cd ${BIGDATA_HOME}**/FusionInsight_ClickHouse_***Version***/**
x_x_**ClickHouseServer/etc**

cat metrika.xml

Step 7 Check whether the cluster topology information on ZooKeeper obtained in **Step 5** is the same as that in the **metrika.xml** file in the component configuration directory in **Step 6**.

- If yes, check whether the alarm is cleared. If the alarm persists, go to **Step 9**.
- If no, go to **Step 8**.

Step 8 On FusionInsight Manager, choose **Cluster** > **Services** > **ClickHouse**, click **More**, and select **Synchronize Configuration**. Then, check whether the service status is normal and whether the alarm is cleared 5 minutes later.

- If yes, no further action is required.
- If no, go to **Step 9**.

Collect the fault information.

Step 9 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 10 Expand the **Service** drop-down list, and select **ClickHouse** for the target cluster.

Step 11 Choose the corresponding host form the host list.

Step 12 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 13 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.240 ALM-45426 ClickHouse Service Quantity Quota Usage in ZooKeeper Exceeds the Threshold

Description

The alarm module checks the quota usage of the ClickHouse service in the ZooKeeper every 60 seconds. This alarm is generated when the alarm module detects that the usage exceeds the threshold (90%).

This alarm is cleared when the system detects that the usage is lower than the threshold and the alarm is cleared.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45426	Major (default)	Yes

Parameters

Name	Meaning
Source	Specifies the cluster or system for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

After the ZooKeeper quantity quota of the ClickHouse service exceeds the threshold, you cannot perform cluster operations on the ClickHouse service on FusionInsight Manager. As a result, the ClickHouse service cannot be used.

Possible Causes

- When table data is created, inserted, or deleted, the ClickHouse creates znodes on ZooKeeper nodes. As the service volume increases, the number of znodes may exceed the configured threshold.
- No quota limit is set for the metadata directory **/clickhouse** of ClickHouse in ZooKeeper.

Procedure

Check the number of znodes created by ClickHouse on ZooKeeper.

Step 1 Log in to the host where the ZooKeeper client is located and log in to the ZooKeeper client.

Switch to the client installation directory.

Example: **cd /opt/client**

Run the following command to configure environment variables:

source bigdata_env

Run the following command to authenticate the user (skip this step in common mode):

kinit *Component service user*

Run the following command to log in to the client tool:

zkCli.sh -server *service IP address of the node where the ZooKeeper role instance locates*:*client port*

Step 2 Run the following command to check the quota used by the ClickHouse in the ZooKeeper and check whether the quota information is correctly set:

listquota /clickhouse

```
absolute path is /zookeeper/quota/clickhouse
Quota for path /clickhouse does not exist.
```

If the preceding information indicates that the quota configuration is incorrect, go to **Step 3**.

If no, go to **Step 5**.

Step 3 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **ZooKeeper**. On the displayed page, click **Configurations** and click **All Configurations**. On this sub-tab page, search for **quotas.auto.check.enable** to check whether its value is **true**.

If the value is not **true**, change the value to **true** and click **Save**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **ClickHouse**, click **More**, and select **Synchronize Configuration**. After the synchronization is successful, go to **Step 1**.

Step 5 Run the following command and check whether the ratio of the **count** value of **Output stat** to the **count** value of **Output quota** in the command output is greater than **0.9**:

listquota /clickhouse

```
absolute path is /zookeeper/quota/clickhouse
Output quota for /clickhouse count=200000,bytes=1000000000
Output stat for /clickhouse count=2667,bytes=60063
```

In the preceding information, the **count** value of **Output stat** is **2667**, and the **count** value of **Output quota** is **200000**.

- If yes, go to **Step 6**.
- If no, check whether the alarm is cleared 5 minutes later. If the alarm persists, go to **Step 8**.

Step 6 On FusionInsight Manager, choose **Cluster** > **Services** > **ClickHouse** > **Configurations** > **All Configurations**, search for the **clickhouse.zookeeper.quota.node.count** parameter, and change the value of this parameter to twice the **count** value of **Output stat** in **Step 5**.

Step 7 Restart the ClickHouse instance for which the alarm is generated, and check whether the alarm is cleared 5 minutes later.

- If yes, no further action is required.
- If no, perform **Step 6** again, and check whether the alarm is cleared 5 minutes later. If the alarm persists, go to **Step 8**.

Collect the fault information.

Step 8 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 9 Expand the **Service** drop-down list, and select **ClickHouse** for the target cluster.

Step 10 Choose the corresponding host form the host list.

Step 11 Click ✏ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.241 ALM-45427 ClickHouse Service Capacity Quota Usage in ZooKeeper Exceeds the Threshold

Description

The alarm module checks the quota usage of the ClickHouse service in the ZooKeeper every 60 seconds. This alarm is generated when the alarm module detects that the usage exceeds the threshold (90%).

This alarm is cleared when the system detects that the usage is lower than the threshold and the alarm is cleared.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45427	Major (default)	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

After the ZooKeeper quantity quota of the ClickHouse service exceeds the threshold, you cannot perform cluster operations on the ClickHouse service on FusionInsight Manager. As a result, the ClickHouse service cannot be used.

Possible Causes

- When table data is created, inserted, or deleted, the ClickHouse creates znodes on ZooKeeper nodes. As the service volume increases, the capacity of znodes may exceed the configured threshold.
- No quota limit is set for the metadata directory **/clickhouse** of ClickHouse in ZooKeeper.

Procedure

Check the znode capacity of the ClickHouse in the ZooKeeper.

Step 1 Log in to the host where the ZooKeeper client is located and log in to the ZooKeeper client.

Switch to the client installation directory.

Example: **cd /opt/client**

Run the following command to configure environment variables:

source bigdata_env

Run the following command to authenticate the user (skip this step in common mode):

kinit *Component service user*

Run the following command to log in to the client tool:

zkCli.sh -server *service IP address of the node where the ZooKeeper role instance locates*:*client port*

Step 2 Run the following command to check the quota used by the ClickHouse in the ZooKeeper and check whether the quota information is correctly set:

listquota /clickhouse

```
absolute path is /zookeeper/quota/clickhouse
Quota for path /clickhouse does not exist.
```

- If the preceding information indicates that the quota configuration is incorrect, go to **Step 3**.
- If not, go to **Step 5**.

Step 3 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **ZooKeeper**. On the displayed page, click **Configurations** and click **All Configurations**. On this sub-tab page, search for **quotas.auto.check.enable** to check whether its value is **true**.

If the value is not **true**, change the value to **true** and click **Save**.

Step 4 On FusionInsight Manager, choose **Cluster** > **Services** > **ClickHouse**, click **More**, and select **Synchronize Configuration**. After the synchronization is successful, go to **Step 1**.

Step 5 Run the following command and check whether the ratio of the **bytes** value of **Output stat** to the **bytes** value of **Output quota** in the command output is greater than **0.9**:

listquota /clickhouse

```
absolute path is /zookeeper/quota/clickhouse
Output quota for /clickhouse count=200000,bytes=1000000000
Output stat for /clickhouse count=2667,bytes=60063
```

In the preceding information, the **bytes** value of **Output stat** is **60063**, and the **bytes** value of **Output quota** is **1000000000**.

- If yes, go to **Step 6**.
- If no, check whether the alarm is cleared 5 minutes later. If the alarm persists, go to **Step 8**.

Step 6 On FusionInsight Manager, choose **Cluster** > **Services** > **ClickHouse** > **Configurations** > **All Configurations**, search for the **clickhouse.zookeeper.quota.size** parameter, and change the value of this parameter to twice the **bytes** value of **Output stat** in **Step 5**.

Step 7 Restart the ClickHouse instance for which the alarm is generated, and check whether the alarm is cleared 5 minutes later.

- If yes, no further action is required.
- If no, perform **Step 6** again, and check whether the alarm is cleared 5 minutes later. If the alarm persists, go to **Step 8**.

Collect the fault information.

Step 8 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 9 Expand the **Service** drop-down list, and select **ClickHouse** for the target cluster.

Step 10 Choose the corresponding host form the host list.

Step 11 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 12 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm is automatically cleared after the fault is rectified.

Related Information

None

9.13.242 ALM-45428 ClickHouse Disk I/O Exception

Description

This alarm is generated when the alarm module detects EIO or EROFS errors during ClickHouse read and write every 60 seconds.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45428	Major (default)	No

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

- ClickHouse fails to read and write data. The INSERT, SELECT, and CREATE operations on the local tables may be abnormal. Distributed tables are not affected.
- Services are affected, and I/Os fail.

Possible Causes

The disk is aged or has bad sectors.

Procedure

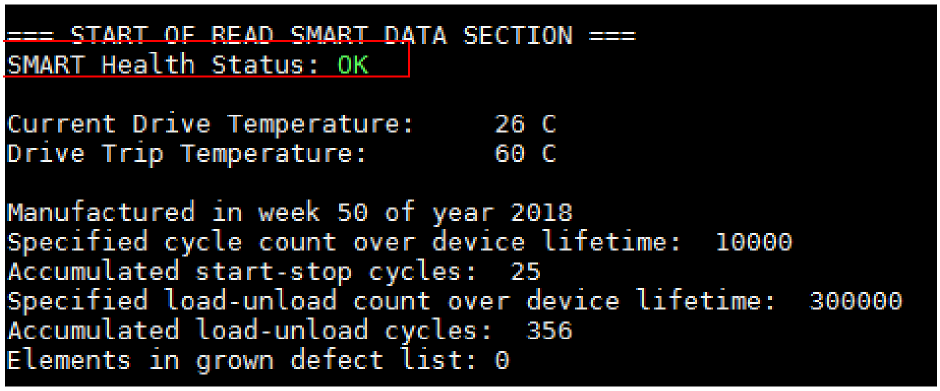
Step 1 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms** > **ALM-45428 ClickHouse Disk I/O Exception**. Check the role name and the IP address of the host where the alarm is generated in **Location**.

Step 2 Use PuTTY to log in to the node for which the fault is generated as user **root**.

Step 3 Run the **df -h** command to check the mount directory and find the disk mounted to the faulty directory.

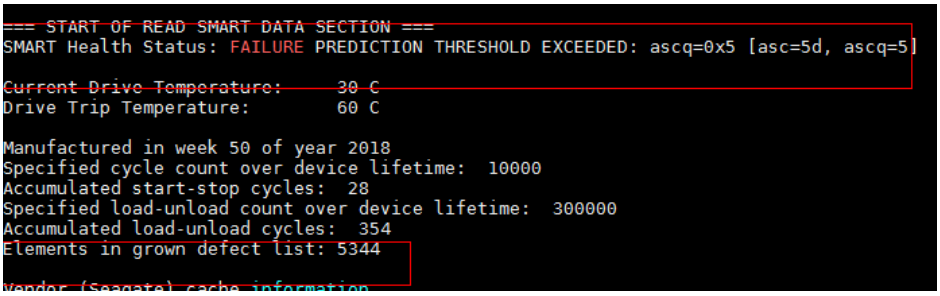
Step 4 Run the **smartctl -a /dev/sd*** command to check disks.

- If **SMART Health Status: OK** is displayed, as shown in the following figure, the disk is healthy. In this case, go to **Step 7**.



```
=== START OF READ SMART DATA SECTION ===
SMART Health Status: OK

Current Drive Temperature:     26 C
Drive Trip Temperature:        60 C

Manufactured in week 50 of year 2018
Specified cycle count over device lifetime:  10000
Accumulated start-stop cycles:  25
Specified load-unload count over device lifetime:  300000
Accumulated load-unload cycles:  356
Elements in grown defect list: 0
```

- If the number following **Elements in grown defect list** is not 0, as shown in the following figure, the disk may have bad sectors. If **SMART Health Status: FAILURE** is displayed, the disk is in the sub-health state. In this case, go to **Step 5**.



```
=== START OF READ SMART DATA SECTION ===
SMART Health Status: FAILURE PREDICTION THRESHOLD EXCEEDED: ascq=0x5 [asc=5d, ascq=5]

Current Drive Temperature:     30 C
Drive Trip Temperature:        60 C

Manufactured in week 50 of year 2018
Specified cycle count over device lifetime:  10000
Accumulated start-stop cycles:  28
Specified load-unload count over device lifetime:  300000
Accumulated load-unload cycles:  354
Elements in grown defect list: 5344

Vendor (Seagate) cache information
```

Step 5 Rectify the fault by following the instructions provided in "Hard Disk Mounted to the ClickHouse Partition Directory Is Faulty" in .

Step 6 After the fault is rectified, manually clear the alarm on FusionInsight Manager and check whether the alarm is generated again during the periodic check.

- If yes, go to **Step 7**.
- If no, no further action is required.

Collect the fault information.

Step 7 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 8 Expand the **Service** drop-down list, and select **ClickHouse** for the target cluster.

Step 9 Choose the corresponding host form the host list.

Step 10 Click in the upper right corner, and set **Start Date** and **End Date** for log collection to 1 hour ahead of and after the alarm generation time, respectively. Then, click **Download**.

Step 11 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

If the alarm has no impact, manually clear the alarm.

Related Information

None

9.13.243 ALM-45429 Table Metadata Synchronization Failed on the Added ClickHouse Node

NOTE

This section applies only to MRS 3.1.2 or later.

Description

This alarm is generated when the local table corresponding to the distributed table fails to be created during ClickHouse capacity expansion.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45429	Major	No

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The distributed table fails to be queried.

Possible Causes

A node is stopped or faulty during capacity expansion.

Procedure

Step 1 On FusionInsight Manager, choose **Cluster** > **Services** > **ClickHouse** > **Instance**.

Step 2 Check whether an instance is stopped, decommissioned, or faulty.

- If yes, go to **Step 3**.
- If no, go to **Step 4**.

Step 3 Start the instance or rectify the instance fault until all instances are running properly.

Step 4 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**, locate this alarm and the faulty host based on the location information.

Step 5 Log in to the faulty host as user **omm**.

Step 6 Run the following commands to initialize environment variables:

source *Cluster installation directory*/**FusionInsight_ClickHouse_*/*_*_ClickHouseServer/etc/ENV_VARS**

source *Cluster installation directory*/**FusionInsight_ClickHouse_*/*_*_ClickHouseServer/etc/clickhouse-env.sh**

export CLICKHOUSE_CONF_DIR=${CLICKHOUSE_CONF_DIR}

Step 7 Run the following command to run the metadata synchronization tool to synchronize metadata from the existing node to the faulty node:

sh *Cluster installation directory*/**FusionInsight_ClickHouse_*/install/FusionInsight-ClickHouse-*/clickhouse/sbin/clickhouse-create-meta.sh true**

Step 8 Run the following command to view the log information and check whether the metadata has been synchronized:

vim /var/log/Bigdata/clickhouse/clickhouseServer/start.log

- If the synchronization is complete, go to **Step 9**.
- If the synchronization fails, go to **Step 10**.

Step 9 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. In the **Alarm ID** column, locate the corresponding alarm and click **Clear** in the **Operation** column. In the displayed dialog box, click **OK** to manually clear the alarm.

Collect the fault information.

Step 10 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 11 Expand the **Service** drop-down list, select **ClickHouse** for the target cluster, and click **OK**.

Step 12 Choose the corresponding host form the host list.

Step 13 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 14 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm needs to be manually cleared after the fault is rectified.

Related Information

None

9.13.244 ALM-45430 Permission Metadata Synchronization Failed on the Added ClickHouse Node

NOTE

This section applies only to MRS 3.1.2 or later.

Description

This alarm is generated when user and permission information fails to be synchronized during ClickHouse capacity expansion.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45430	Major	No

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
RoleName	Specifies the role for which the alarm is generated.
HostName	Specifies the host for which the alarm is generated.

Impact on the System

The created user does not have operation permissions on the node.

Possible Causes

A node is stopped or faulty during capacity expansion.

Procedure

Step 1 On FusionInsight Manager, choose **Cluster** > **Services** > **ClickHouse** > **Instance**.

Step 2 Check whether an instance is stopped, decommissioned, or faulty.

- If yes, go to **Step 3**.
- If no, go to **Step 4**.

Step 3 Start the instance or rectify the instance fault until all instances are running properly.

Step 4 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**, locate this alarm and the faulty host based on the location information.

Step 5 Log in to the faulty host as user **omm**.

Step 6 Run the following commands to initialize environment variables:

source *Cluster installation directory*/**FusionInsight_ClickHouse_*/*_*_ClickHouseServer/etc/ENV_VARS**

source *Cluster installation directory*/**FusionInsight_ClickHouse_*/*_*_ClickHouseServer/etc/clickhouse-env.sh**

export CLICKHOUSE_CONF_DIR=${CLICKHOUSE_CONF_DIR}

Step 7 Run the following command to run the metadata synchronization tool to synchronize metadata from the existing node to the faulty node:

sh *Cluster installation directory*/**FusionInsight_ClickHouse_*/install/FusionInsight-ClickHouse-*/clickhouse/sbin/clickhouse-create-meta.sh true**

Step 8 Run the following command to view the log information and check whether the metadata has been synchronized:

vim /var/log/Bigdata/clickhouse/clickhouseServer/start.log

If the synchronization is complete, go to **Step 9**.

If the synchronization fails, go to **Step 10**.

Step 9 On FusionInsight Manager, choose **O&M** > **Alarm** > **Alarms**. In the **Alarm ID** column, locate the corresponding alarm and click **Clear** in the **Operation** column. In the displayed dialog box, click **OK** to manually clear the alarm.

Collect the fault information.

Step 10 On FusionInsight Manager, choose **O&M**. In the navigation pane on the left, choose **Log** > **Download**.

Step 11 Expand the **Service** drop-down list, select **ClickHouse** for the target cluster, and click **OK**.

Step 12 Choose the corresponding host form the host list.

Step 13 Click ✎ in the upper right corner, and set **Start Date** and **End Date** for log collection to 30 minutes ahead of and after the alarm generation time respectively. Then, click **Download**.

Step 14 Contact O&M personnel and provide the collected logs.

----End

Alarm Clearing

This alarm needs to be manually cleared after the fault is rectified.

Related Information

None

9.13.245 ALM-45635 FlinkServer Job Execution Failure

This section applies to MRS 3.1.2 or later.

Description

The system checks whether FlinkServer jobs fail to be executed every 10 seconds. This alarm is generated when a FlinkServer job fails. This alarm is cleared when the job is successfully restarted.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45635	Major	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
JobName	Specifies the job for which the alarm is generated.

Impact on the System

This alarm has no impact on the system.

Possible Causes

You can view failure causes in specific logs.

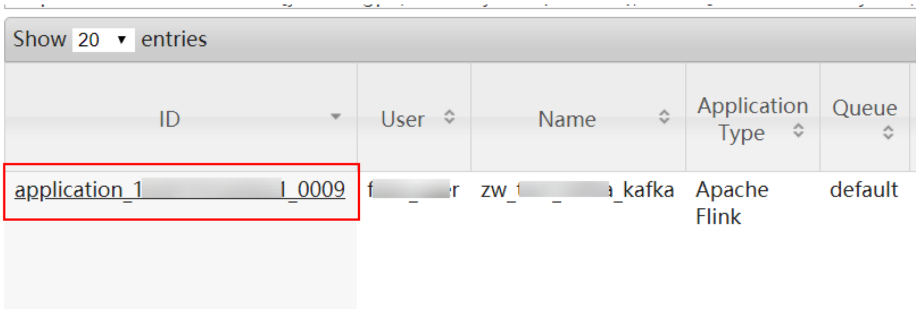
Procedure

Step 1 Log in to Manager as a user who has the FlinkServer management permission.

Step 2 Choose **Cluster** > **Services** > **Yarn** and click the link next to **ResourceManager WebUI** to go to the Yarn page.

Step 3 Locate the failed job based on its name displayed in **Location**, search for and record the application ID of the failed job, and check whether the job logs are available on the Yarn page.

Figure 9-35 Application ID of a job



Show 20 entries

ID	User	Name	Application Type	Queue
application_1[redacted]_0009	f[redacted]r	zw_[redacted]_kafka	Apache Flink	default

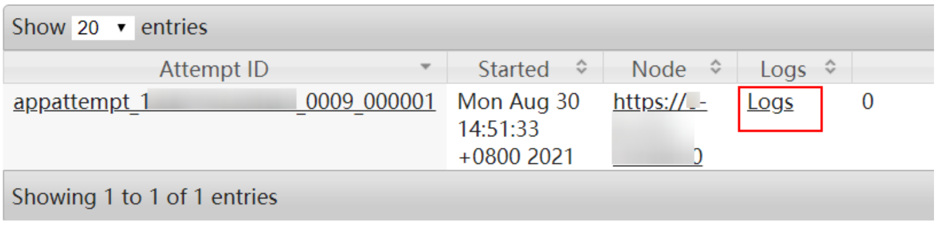
If yes, go to **Step 4**.

If no, go to **Step 6**.

Step 4 Click the application ID of the failed job to go to the job page.

1. Click **Logs** in the **Logs** column to view JobManager logs.

Figure 9-36 Clicking Logs



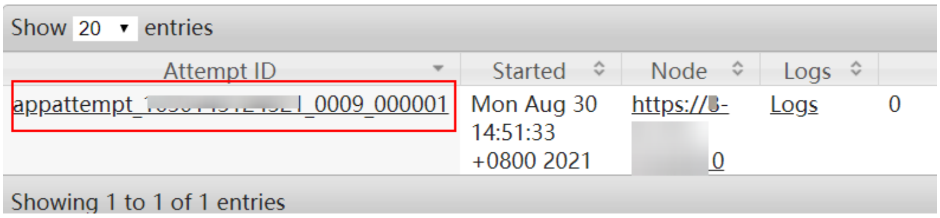
Show 20 ▾ entries

Attempt ID	Started	Node	Logs	
appattempt_1[redacted]_0009_000001	Mon Aug 30 14:51:33 +0800 2021	https://[redacted]	Logs	0

Showing 1 to 1 of 1 entries

2. Click the ID in the **Attempt ID** column and click **Logs** in the **Logs** column to view TaskManager logs.

Figure 9-37 Clicking the ID in the Attempt ID column

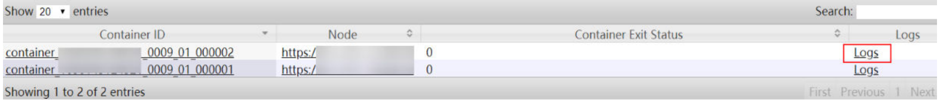


Show 20 ▾ entries

Attempt ID	Started	Node	Logs	
appattempt_[redacted]_0009_000001	Mon Aug 30 14:51:33 +0800 2021	https://[redacted]	Logs	0

Showing 1 to 1 of 1 entries

Figure 9-38 Clicking Logs



Show 20 ▾ entries Search:

Container ID	Node	Container Exit Status	Logs
container_[redacted]_0009_01_000002	https:/[redacted]	0	Logs
container_[redacted]_0009_01_000001	https:/[redacted]	0	Logs

Showing 1 to 2 of 2 entries

NOTE

You can also log in to Manager as a user who has the FlinkServer management permission, choose **Cluster** > **Services** > **Flink**, click the link next to **Flink WebUI**. On the displayed Flink web UI, click **Job Management** and choose **More** > **Job Monitoring** in the **Operation** column to view the TaskManager logs.

Step 5 View the logs of the failed job to rectify the fault, or contact the O&M personnel personnel and send the collected fault logs. No further action is required.

If logs are unavailable on the Yarn page, download logs from HDFS.

Step 6 On Manager, choose **Cluster** > **Services** > **HDFS**, click the link next to **NameNode WebUI** to go to the HDFS page, select **Utilities** > **Browse the file system**, and download logs in the **/tmp/logs/***User name***/logs/***Application ID of the failed job* directory.

Step 7 View the logs of the failed job to rectify the fault, or contact the O&M personnel personnel and send the collected fault logs.

----End

Alarm Clearing

After the job is successfully restarted, the alarm is cleared if it has been reported.

Related Information

None

9.13.246 ALM-45636 FlinkServer Job Checkpoints Keep Failing

This section applies to MRS 3.1.2 or later.

Description

The system checks the number of consecutive checkpoint failures based on the configured alarm checking interval. This alarm is generated when the number of consecutive checkpoint failures of a FlinkServer job reaches the configured threshold. This alarm is cleared when checkpoints are recovered or the job is successfully restarted.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45636	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
JobName	Specifies the job for which the alarm is generated.

Impact on the System

This alarm has no impact on the system.

Possible Causes

You can view failure causes in specific logs.

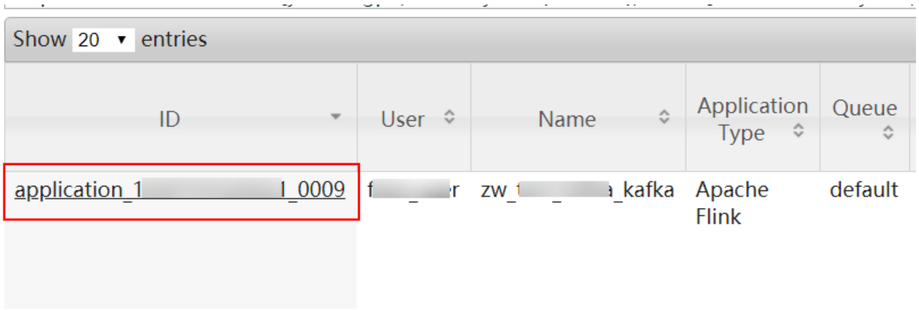
Procedure

Step 1 Log in to Manager as a user who has the FlinkServer management permission.

Step 2 Choose **Cluster** > **Services** > **Yarn** and click the link next to **ResourceManager WebUI** to go to the Yarn page.

Step 3 Locate the failed job based on its name displayed in **Location**, search for and record the application ID of the failed job, and check whether the job logs are available on the Yarn page.

Figure 9-39 Application ID of a job



Show 20 entries

ID	User	Name	Application Type	Queue
application_1[redacted]l_0009	f[redacted]r	zw_[redacted]_kafka	Apache Flink	default

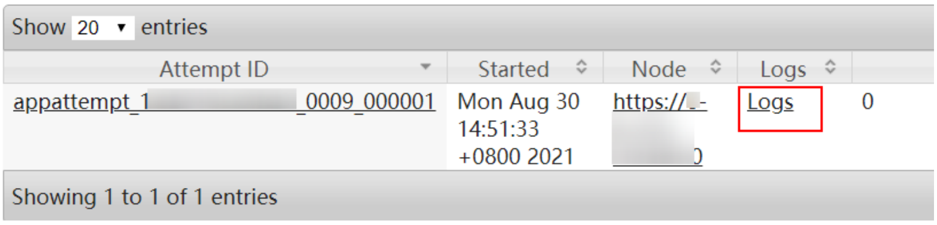
If yes, go to **Step 4**.

If no, go to **Step 6**.

Step 4 Click the application ID of the failed job to go to the job page.

1. Click **Logs** in the **Logs** column to view JobManager logs.

Figure 9-40 Clicking Logs



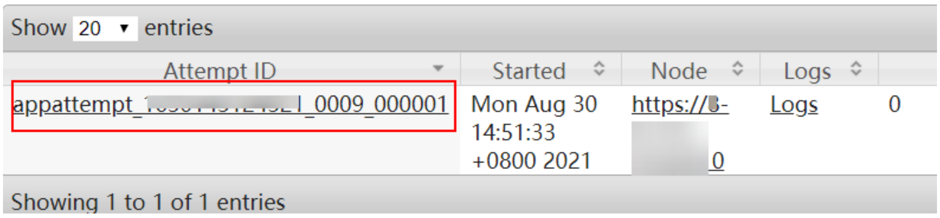
Show 20 entries

Attempt ID	Started	Node	Logs	
appattempt_1[redacted]_0009_000001	Mon Aug 30 14:51:33 +0800 2021	https://[redacted]0	Logs	0

Showing 1 to 1 of 1 entries

2. Click the ID in the **Attempt ID** column and click **Logs** in the **Logs** column to view TaskManager logs.

Figure 9-41 Clicking the ID in the Attempt ID column

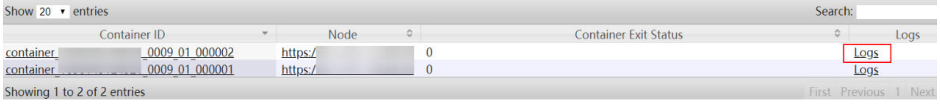


Show 20 entries

Attempt ID	Started	Node	Logs	
appattempt_1[redacted]l_0009_000001	Mon Aug 30 14:51:33 +0800 2021	https://[redacted]0	Logs	0

Showing 1 to 1 of 1 entries

Figure 9-42 Clicking Logs



Show 20 entries Search:

Container ID	Node	Container Exit Status	Logs
container_[redacted]_0009_01_000002	https:/[redacted]	0	Logs
container_[redacted]_0009_01_000001	https:/[redacted]	0	Logs

Showing 1 to 2 of 2 entries First Previous 1 Next

NOTE

You can also log in to Manager as a user who has the FlinkServer management permission, choose **Cluster** > **Services** > **Flink**, click the link next to **Flink WebUI**. On the displayed Flink web UI, click **Job Management** and choose **More** > **Job Monitoring** in the **Operation** column to view the TaskManager logs.

Step 5 View the logs of the failed job to rectify the fault, or contact the O&M personnel personnel and send the collected fault logs. No further action is required.

If logs are unavailable on the Yarn page, download logs from HDFS.

Step 6 On Manager, choose **Cluster** > **Services** > **HDFS**, click the link next to **NameNode WebUI** to go to the HDFS page, select **Utilities** > **Browse the file system**, and download logs in the **/tmp/logs/***User name***/logs/***Application ID of the failed job* directory.

Step 7 View the logs of the failed job to rectify the fault, or contact the O&M personnel personnel and send the collected fault logs.

----End

Alarm Clearing

This alarm is cleared when FlinkServer job checkpoints are recovered or the job is successfully restarted.

Related Information

None

9.13.247 ALM-45637 FlinkServer Task Is Continuously Under Back Pressure

This section applies to MRS 3.1.2 or later.

Description

The system checks the back pressure duration of FlinkServer tasks based on the configured alarm checking interval. This alarm is generated when the back pressure duration of a FlinkServer task reaches the configured threshold. This alarm is cleared when the task back pressure is recovered or the job is successfully restarted.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45637	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
JobName	Specifies the job for which the alarm is generated.

Impact on the System

This alarm has no impact on the system.

Possible Causes

You can view the causes in the specific logs.

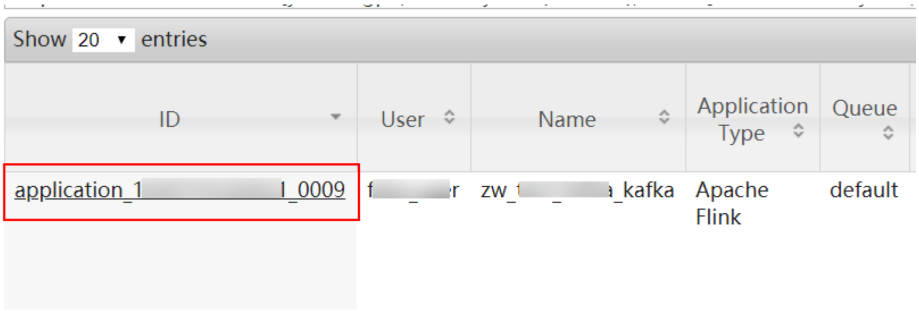
Procedure

Step 1 Log in to Manager as a user who has the FlinkServer management permission.

Step 2 Choose **Cluster** > **Services** > **Yarn** and click the link next to **ResourceManager WebUI** to go to the Yarn page.

Step 3 Locate the failed job based on its name displayed in **Location**, search for and record the application ID of the failed job, and check whether the job logs are available on the Yarn page.

Figure 9-43 Application ID of a job



Show 20 ▾ entries

ID	User	Name	Application Type	Queue
application_1[redacted]_0009	f[redacted]r	zw_[redacted]_kafka	Apache Flink	default

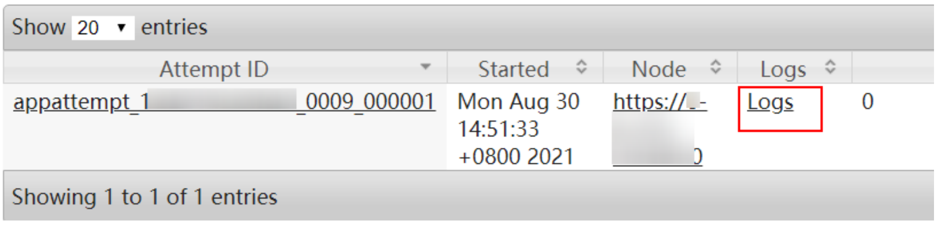
If yes, go to **Step 4**.

If no, go to **Step 6**.

Step 4 Click the application ID of the failed job to go to the job page.

1. Click **Logs** in the **Logs** column to view JobManager logs.

Figure 9-44 Clicking Logs



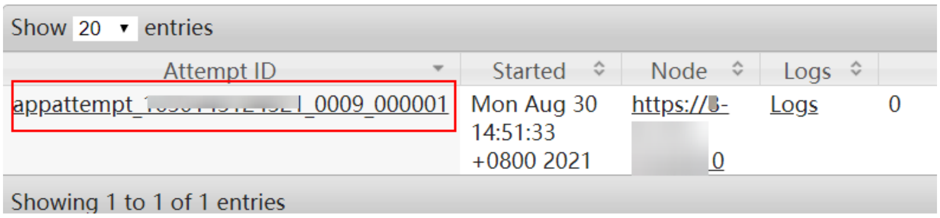
Show 20 ▾ entries

Attempt ID	Started	Node	Logs	
appattempt_1[redacted]_0009_000001	Mon Aug 30 14:51:33 +0800 2021	https://[redacted]0	Logs	0

Showing 1 to 1 of 1 entries

2. Click the ID in the **Attempt ID** column and click **Logs** in the **Logs** column to view TaskManager logs.

Figure 9-45 Clicking the ID in the Attempt ID column

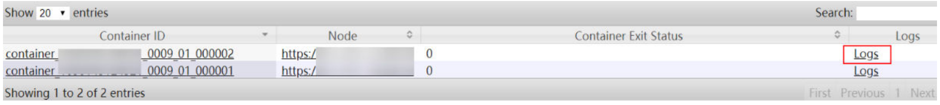


Show 20 ▾ entries

Attempt ID	Started	Node	Logs	
appattempt_[redacted]_0009_000001	Mon Aug 30 14:51:33 +0800 2021	https://[redacted]0	Logs	0

Showing 1 to 1 of 1 entries

Figure 9-46 Clicking Logs



Show 20 ▾ entries Search:

Container ID	Node	Container Exit Status	Logs
container_[redacted]_0009_01_000002	https:/[redacted]	0	Logs
container_[redacted]_0009_01_000001	https:/[redacted]	0	Logs

Showing 1 to 2 of 2 entries

NOTE

You can also log in to Manager as a user who has the FlinkServer management permission, choose **Cluster** > **Services** > **Flink**, click the link next to **Flink WebUI**. On the displayed Flink web UI, click **Job Management** and choose **More** > **Job Monitoring** in the **Operation** column to view the TaskManager logs.

Step 5 View the logs of the failed job to rectify the fault, or contact the O&M personnel personnel and send the collected fault logs. No further action is required.

If logs are unavailable on the Yarn page, download logs from HDFS.

Step 6 On Manager, choose **Cluster** > **Services** > **HDFS**, click the link next to **NameNode WebUI** to go to the HDFS page, select **Utilities** > **Browse the file system**, and download logs in the **/tmp/logs/***User name***/logs/***Application ID of the failed job* directory.

Step 7 View the logs of the failed job to rectify the fault, or contact the O&M personnel personnel and send the collected fault logs.

----End

Alarm Clearing

This alarm is cleared when FlinkServer task back pressure is recovered or the job is successfully restarted.

Related Information

None

9.13.248 ALM-45638 Number of Restarts After FlinkServer Job Failures Exceeds the Threshold

This section applies to MRS 3.1.2 or later.

Description

The system checks the number of FlinkServer job restarts based on the alarm checking interval. This alarm is generated when the number exceeds the configured threshold. This alarm is cleared when the job is restarted.

Attribute

Alarm ID	Alarm Severity	Auto Clear
45638	Minor	Yes

Parameters

Name	Meaning
Source	Specifies the cluster for which the alarm is generated.
ServiceName	Specifies the service for which the alarm is generated.
JobName	Specifies the job for which the alarm is generated.

Impact on the System

This alarm has no impact on the system.

Possible Causes

You can view the causes in the specific logs.

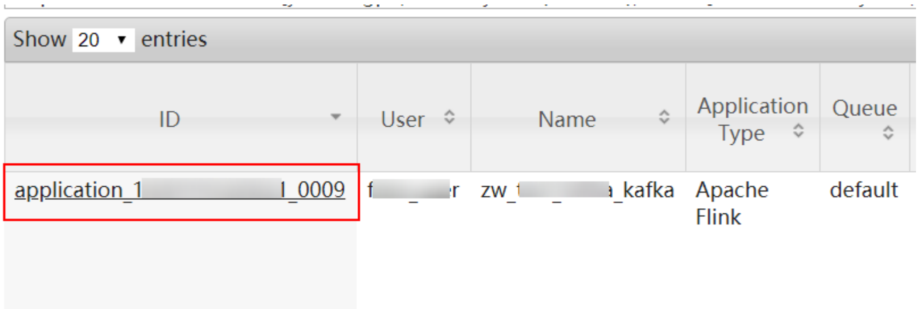
Procedure

Step 1 Log in to Manager as a user who has the FlinkServer management permission.

Step 2 Choose **Cluster** > **Services** > **Yarn** and click the link next to **ResourceManager WebUI** to go to the Yarn page.

Step 3 Locate the failed job based on its name displayed in **Location**, search for and record the application ID of the failed job, and check whether the job logs are available on the Yarn page.

Figure 9-47 Application ID of a job



Show 20 entries

ID	User	Name	Application Type	Queue
application_1[redacted]_0009	f[redacted]r	zw_[redacted]_kafka	Apache Flink	default

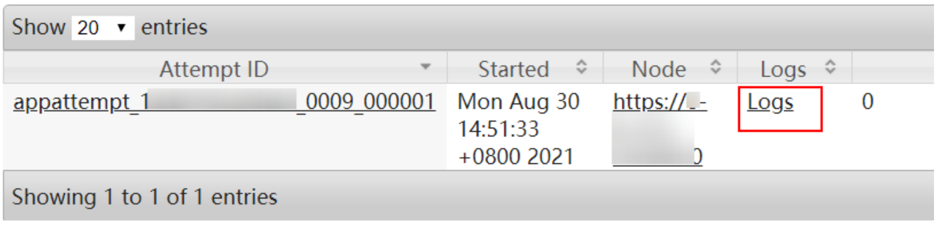
If yes, go to **Step 4**.

If no, go to **Step 6**.

Step 4 Click the application ID of the failed job to go to the job page.

1. Click **Logs** in the **Logs** column to view JobManager logs.

Figure 9-48 Clicking Logs



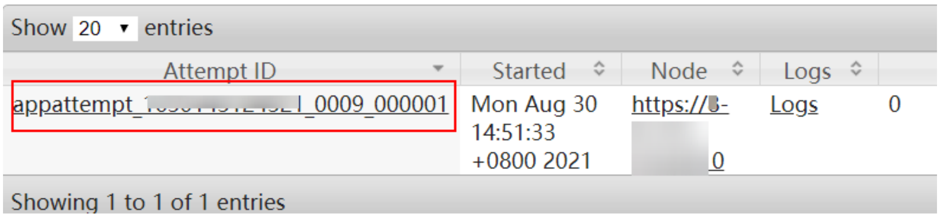
Show 20 ▾ entries

Attempt ID	Started	Node	Logs	
appattempt_1[redacted]_0009_000001	Mon Aug 30 14:51:33 +0800 2021	https://[redacted]0	Logs	0

Showing 1 to 1 of 1 entries

2. Click the ID in the **Attempt ID** column and click **Logs** in the **Logs** column to view TaskManager logs.

Figure 9-49 Clicking the ID in the Attempt ID column

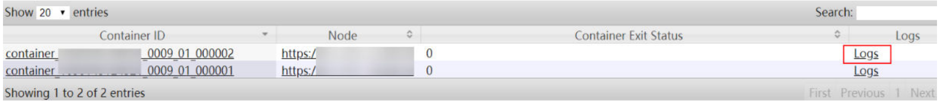


Show 20 ▾ entries

Attempt ID	Started	Node	Logs	
appattempt_1[redacted]1_0009_000001	Mon Aug 30 14:51:33 +0800 2021	https://[redacted]0	Logs	0

Showing 1 to 1 of 1 entries

Figure 9-50 Clicking Logs



Show 20 ▾ entries Search:

Container ID	Node	Container Exit Status	Logs
container_[redacted]_0009_01_000002	https:/[redacted]	0	Logs
container_[redacted]_0009_01_000001	https:/[redacted]	0	Logs

Showing 1 to 2 of 2 entries First Previous 1 Next

NOTE

You can also log in to Manager as a user who has the FlinkServer management permission, choose **Cluster** > **Services** > **Flink**, click the link next to **Flink WebUI**. On the displayed Flink web UI, click **Job Management** and choose **More** > **Job Monitoring** in the **Operation** column to view the TaskManager logs.

Step 5 View the logs of the failed job to rectify the fault, or contact the O&M personnel personnel and send the collected fault logs. No further action is required.

If logs are unavailable on the Yarn page, download logs from HDFS.

Step 6 On Manager, choose **Cluster** > **Services** > **HDFS**, click the link next to **NameNode WebUI** to go to the HDFS page, select **Utilities** > **Browse the file system**, and download logs in the **/tmp/logs/***User name***/logs/***Application ID of the failed job* directory.

Step 7 View the logs of the failed job to rectify the fault, or contact the O&M personnel personnel and send the collected fault logs.

----End

Alarm Clearing

This alarm is cleared when the FlinkServer job is successfully restarted.

Related Information

None

10 MRS Manager Operation Guide (Applicable to 2.x and Earlier Versions)

10.1 Introduction to MRS Manager

Overview

MRS manages and analyzes massive data and helps you rapidly obtain desired data from structured and unstructured data. The structure of open-source components is complex. The installation, configuration, and management processes are time- and labor-consuming. MRS Manager is a unified enterprise-level cluster management platform and provides the following functions:

- Cluster monitoring enables you to quickly view the health status of hosts and services.
- Graphical metric monitoring and customization enable you to quickly obtain key information about the system.
- Service property configurations can meet service performance requirements.
- With cluster, service, and role instance functions, you can start or stop services and clusters in one click.

Introduction to the MRS Manager GUI

MRS Manager provides a unified cluster management platform, facilitating rapid and easy O&M for clusters. For details about how to access MRS Manager, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Table 10-1 describes the functions of each operation entry.

Table 10-1 Functions of each entry on the operation bar

Parameter	Function
Dashboard	Displays the status of all services, main monitoring indicators of each service, and host status in charts, such as bar charts, line charts, and tables. You can customize a dashboard for the key monitoring indicators and drag it to any position on the interface. The system dashboard page supports automatic data update.
Services	Provides the service monitoring, operation, and configuration guidance, which helps you manage services in a unified manner.
Hosts	Provides guidance on how to monitor, operate, and configure hosts, helping you manage hosts in a unified manner.
Alarms	Supports alarm query and provides guidance on alarm handling, helping you identify and rectify product faults and potential risks in a timely manner to ensure normal system operation.
Audit	Allows authorized users to query and export audit logs, helping you to view all user activities and operations.
Tenant	Provides a unified tenant management platform.
System	Provides monitoring, alarm configuration management, and backup management.

Go to the **System** tab page, and switch to another function pages through shortcuts. See **Table 10-2**.

The following is an example of quick redirection through shortcuts:

Step 1 On MRS Manager, click **System**.

Step 2 On the **System** tab page, click a function link. The function page is displayed.

For example, in the **Backup and Restoration** area, click **Back Up Data**. The page for backing up data is displayed.

Step 3 Move the cursor to the left border of the browser window. The **System** black shortcut menu is displayed. After you move the cursor out of the menu, the menu is collapsed.

Step 4 In the shortcut menu that is displayed, you can click a function link to go to the corresponding function page.

For example, choose **Maintenance > Export Log**. The page for exporting logs is displayed.

----End

Table 10-2 Shortcut menus on the **System** tab page

Menu	Function Link
Backup and Restoration	Back Up Data
	Restore Data
Maintenance	Export Log
	Export Audit Log
	Check Health Status
Monitoring and Alarm	Configure Syslog
	Configure Alarm Threshold
	Configure SNMP
	Configure Monitoring Metric Dump
	Configure Resource Contribution Ranking
Permission	Manage User
	Manage User Group
	Manage Role
	Configure Password Policy
	Change OMS Database Password
Patch	Manage Patch

Reference

MapReduce Service (MRS) is a data analysis service. It is used to manage and analyze massive sets of data.

MRS uses MRS Manager to manage big data components, such as components in the Hadoop ecosystem. Therefore, some concepts on the MRS Console must be different from those on MRS Manager. For details, see **Table 10-3**.

Table 10-3 Difference Comparison

Concept	MRS	MRS Manager
MapReduce Service	Indicates the data analysis cloud service, called MRS. This service includes components such as Hive, Spark, Yarn, HDFS, and ZooKeeper.	Provides a unified management platform for big data components in tenant clusters.

10.2 Checking Running Tasks

Scenario

When you perform operations on MRS Manager to trigger a task, the task execution process and progress are displayed. After the task window is closed, you need to open the task window by using the task management function.

MRS Manager reserves 10 latest tasks by default, for example, restarting services, synchronizing service configurations, and performing health check.

Procedure

Step 1 On MRS Manager, click 📋 to open the task list.

You can view the following information in the task list: **Name**, **Status**, **Progress**, **Start Time** and **End Time**.

Step 2 Click the target task name to view the detailed information about the running task.

----End

10.3 Monitoring Management

10.3.1 Dashboard

On MRS Manager, nodes in a cluster can be classified into management nodes, control nodes, and data nodes. The change trends of key host monitoring metrics on each type of node can be calculated and displayed as curve charts in reports based on the customized periods. If a host belongs to multiple node types, the metric statistics will be repeatedly collected.

This section provides overview of MRS clusters and describes how to view, customize, and export node monitoring metrics on MRS Manager.

Procedure

Step 1 Log in to MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 Choose **Dashboard** on MRS Manager.

Step 3 In **Period**, you can specify a period to view monitoring data. The options are as follows:

- Real time
- Last 3 hours
- Last 6 hours
- Last 24 hours

- Last week
- Last month
- Last 3 months
- Last 6 months
- Customize. If you select this option, you can customize the period for viewing monitoring data.

Step 4 Click **View** to view monitoring data in a period.

- You can view **Health Status** and **Roles** of each service on the **Service Summary** page of MRS Manager.
- Click ? above the curve chart to view details about a metric.

Step 5 Customize a monitoring report.

1. Click **Customize** and select monitoring metrics to be displayed on MRS Manager.

   MRS Manager supports a maximum of 14 monitoring metrics, but at most 12 customized monitoring metrics can be displayed on the page.
   - Cluster Host Health Status
   - Cluster Network Read Speed Statistics
   - Host Network Read Speed Distribution
   - Host Network Write Speed Distribution
   - Cluster Disk Write Speed Statistics
   - Cluster Disk Usage Statistics
   - Cluster Disk Information
   - Host Disk Usage Statistics
   - Cluster Disk Read Speed Statistics
   - Cluster Memory Usage Statistics
   - Host Memory Usage Distribution
   - Cluster Network Write Speed Statistics
   - Host CPU Usage Distribution
   - Cluster CPU Usage Statistics
2. Click **OK** to save the selected monitoring metrics for display.

   NOTE

   Click **Clear** to cancel all the selected monitoring metrics in a batch.

Step 6 Set an automatic refresh interval or click ↻ for an immediate refresh.

The following refresh interval options are supported:

- Refresh every 30 seconds
- Refresh every 60 seconds
- Stop refreshing

  NOTE

  If you select **Full Screen**, the **Dashboard** window will be maximized.

Step 7 Export a monitoring report.

1. Select a period. The options are as follows:
   - Real time
   - Last 3 hours
   - Last 6 hours
   - Last 24 hours
   - Last week
   - Last month
   - Last 3 months
   - Last 6 months
   - Customize. If you select this option, you can customize a time of period to export a report.
2. Click **Export**. MRS Manager will generate a report about the selected monitoring metrics in a specified time of period. Save the report.

   NOTE

   To view the curve charts of monitoring metrics in a specified period, click **View**.

----End

10.3.2 Managing Services and Monitoring Hosts

You can manage the following status and indicators of all services (including role instances) and hosts on the MRS Manager:

- Status information: includes operation, health, configuration, and role instance status.
- Metric information: includes key monitoring metrics for services.
- Metric export: allows you to export monitoring reports.

NOTE

Set an automatic refresh interval or click ↻ for an immediate refresh.

The following refresh interval options are supported:

- Refresh every 30 seconds
- Refresh every 60 seconds
- Stop refreshing

Managing Service Monitoring

Step 1 On MRS Manager, click **Services**.

The service list includes **Service**, **Operating Status**, **Health Status**, **Configuration Status**, **Roles**, and **Operation** are displayed in the component list.

- **Table 10-4** describes the service operating status.

Table 10-4 Service operating status

Status	Description
Started	The service is started.
Stopped	The service is stopped.
Failed to start	Failed to start the role instance.
Failed to stop	Failed to stop the role instance.
Unknown	Indicates initial service status after the background system restarts.

- **Table 10-5** describes the service health status.

Table 10-5 Service health status

Status	Description
Good	Indicates that all role instances in the service are running properly.
Bad	Indicates that the running status of at least one role instance is **Faulty** or the status of the service on which the current service depends is abnormal.
Unknown	Indicates that all role instances in the service are in the **Unknown** state.
Concerning	Indicates that the background system is restarting the service.
Partially Healthy	Indicates that the status of the service on which the service depends is abnormal, and APIs related to the abnormal service cannot be invoked by external systems.

- **Table 10-6** describes the service health status.

Table 10-6 Service configuration status

Status	Description
Synchronized	The latest configuration takes effect.
Expired	The latest configuration does not take effect after the parameter modification. Related services need to be restarted.
Failed	The communication is incorrect or data cannot be read or written during the parameter configuration. Use **Synchronize Configuration** to rectify the fault.

Status	Description
Configuring	Parameters are being configured.
Unknown	Current configuration status cannot be obtained.

By default, the **Service** column is sorted in ascending order. You can click the icon next to **Service**, **Operating Status**, **Health Status**, or **Configuration Status** to change the sorting mode.

Step 2 Click a specified service in the list to view its status and metric information.

Step 3 Customize monitoring metrics and export customized monitoring information.

1. In the **Charts** area, click **Customize** to customize service monitoring metrics.
2. In **Period** area, select a time of period and click **View** to view the monitoring data within the time period.
3. Click **Export** to export the displayed metrics.

----End

Managing Role Instances

Step 1 On MRS Manager, click **Services** and click the target service name in the service list.

Step 2 Click **Instance** to view the role status.

The role instance list contains the **Role**, **Host Name**, **OM IP Address**, **Business IP Address**, **Rack**, **Operation Status**, **Health Status,** and **Configuration Status** of an instance.

- **Table 10-7** shows the configuration status of a role instance.

Table 10-7 Role instance status

Status	Description
Started	The role instance has been started.
Stopped	The role instance has been stopped.
Failed to start	Failed to start the role instance.
Failed to stop	Failed to stop the role instance.
Decommissioning	The role instance is being decommissioned.
Decommissioned	The role instance has been decommissioned.
Recommissioning	The role instance is being recommissioned.
Unknown	Indicates initial role instance status after the background system restarts.

- Table 10-8 shows the health status of a role instance.

Table 10-8 Role instance health status

Status	Description
Good	The role instance is running properly.
Bad	The role instance is abnormal. For example, the port cannot be accessed if PID does not exist.
Unknown	The host where a role instance resides does not connect to the background system.
Concerning	The background system is restarting a role instance.

- Table 10-9 shows the configuration status of a role instance.

Table 10-9 Role instance configuration status

Status	Description
Synchronized	The latest configuration takes effect.
Expired	The latest configuration does not take effect after the parameter modification. Related services need to be restarted.
Failed	The communication is incorrect or data cannot be read or written during the parameter configuration. Use **Synchronize Configuration** to rectify the fault.
Configuring	Parameters are being configured.
Unknown	Current configuration status cannot be obtained.

By default, the **Role** column is sorted in ascending order. You can click the sorting icon next to **Role**, **Host Name**, **OM IP Address**, **Business IP Address**, **Rack**, **Operating Status**, **Health Status**, or **Configuration Status** to change the sorting mode.

You can filter out all instances of the same role in the **Role** column.

You can set search criteria in the role search area by clicking **Advanced Search**, and click **Search** to view specified role information. Click **Reset** to clear the search criteria. Fuzzy search is supported.

Step 3 Click the target role instance to view its status and metric information.

Step 4 Customize monitoring metrics and export customized monitoring information.

1. In the **Charts** area, click **Customize** to customize service monitoring metrics.
2. In **Period** area, select a time of period and click **View** to view the monitoring data within the time period.
3. Click **Export** to export the displayed metrics.

----End

Managing Hosts

Step 1 On MRS Manager, click **Hosts** to view the status of all hosts.

The host list contains the host name, management IP address, service IP address, rack, network speed, operating status, health status, disk usage, memory usage, and CPU usage.

- **Table 10-10** shows the host operating status.

Table 10-10 Host operating status

Status	Description
Normal	The host and service roles on the host are running properly.
Isolated	The host is isolated, and the service roles on the host stop running.

- **Table 10-11** describes the host health status.

Table 10-11 Host health status

Status	Description
Good	The host can properly send heartbeats.
Bad	The host fails to send heartbeats due to timeout.
Unknown	The host initial status is unknown during the operation of adding or deleting a host.

By default, the **Host Name** column is sorted by host name in ascending order. You can click the sorting icon next to **Host Name**, **OM IP Address**, **Business IP Address**, **Rack**, **Network Speed**, **Operating Status**, **Health Status**, **Disk Usage**, **Memory Usage**, or **CPU Usage** to change the sorting mode.

You can set search criteria in the role search area by clicking **Advanced Search**, and click **Search** to view specified role information. Click **Reset** to clear the search criteria. Fuzzy search is supported.

Step 2 Click the target host in the host list to view its status and metric information.

Step 3 Customize monitoring metrics and export customized monitoring information.

1. In the **Charts** area, click **Customize** to customize service monitoring metrics.
2. In **Period** area, select a time of period and click **View** to view the monitoring data within the time period.
3. Click **Export** to export the displayed metrics.

----End

10.3.3 Managing Resource Distribution

On MRS Manager, you can query the top value curves, bottom value curves, or average data curves of key service and host monitoring metrics, that is, the resource distribution information. MRS Manager allows you to view the monitoring data of the last hour.

You can also modify the resource distribution on MRS Manager to display both the top and bottom value curves in service and host resource distribution figures.

Resource distribution of some monitoring metrics is not recorded.

Procedure

- View the resource distribution of service monitoring metrics.
  1. On MRS Manager, click **Services**.
  2. Select the target service from the service list.
  3. Click **Resource Distribution**.

     Select key metrics of the service from **Metric**. MRS Manager displays the resource distribution of the metrics in the last hour.
- View the resource distribution of host monitoring metrics.
  1. Click **Hosts**.
  2. Click the name of the specified host in the host list.
  3. Click **Resource Distribution**.

     Select key metrics of the host from **Metrics**. MRS Manager displays the resource distribution of the metrics in the last hour.
- Configure resource distribution.
  1. On MRS Manager, click **System**.
  2. In **Configuration**, click **Configure Resource Contribution Ranking** under **Monitoring and Alarm**.
  3. Change the number of resources to be displayed.
     - Set **Number of Top Resources** to the number of top values.
     - Set **Number of Bottom Resources** to the number of bottom values.

     NOTE

     The sum of the maximum value and minimum value of resource distribution cannot be greater than 5.
  4. Click **OK** to save the configurations.

     The message "Number of top and bottom resources saved successfully" is displayed in the upper right corner of the page.

10.3.4 Configuring Monitoring Metric Dumping

You can configure interconnection parameters on MRS Manager to save monitoring metric data to a specified FTP server using the FTP or SFTP protocol. In this way, MRS clusters can interconnect with third-party systems. The FTP protocol does not encrypt data, which brings potential security risks. Therefore, the SFTP protocol is recommended.

MRS Manager supports the collection of all the monitoring metric data in the managed clusters. The collection period is 30 seconds, 60 seconds, or 300 seconds. The monitoring metric data is stored to different monitoring files on the FTP server by collection period. The monitoring file naming rule is in the "*Cluster name*_**metric**_*Monitoring metric data collection period*_*File saving time*.**log**" format.

Prerequisites

The ECS corresponding to the dump server must be in the same VPC as the Master node of the MRS cluster, and the Master node can access the IP address and specified port of the dump server. The FTP service on the dump server is running properly.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 In **Configuration**, click **Configure Monitoring Metric Dump** under **Monitoring and Alarm**.

Step 3 Table 10-12 describes dump parameters.

Table 10-12 Dump parameters

Parameter	Description
FTP IP Address	Mandatory. This parameter specifies the FTP server for storing monitoring files after the monitoring indicator data is interconnected.
FTP Port	Mandatory. This parameter specifies the port connected to the FTP server.
FTP Username	Mandatory. This parameter specifies the username for logging in to the FTP server.
FTP Password	Mandatory. This parameter specifies the password for logging in to the FTP server.
Save Path	Mandatory. This parameter specifies the path for storing monitoring files on the FTP server.
Dump Interval (s)	Mandatory. This parameter specifies the interval at which monitoring files are periodically stored on the FTP server, in seconds.
Dump Mode	Mandatory. This parameter specifies the protocol used for sending monitoring files. This parameter is mandatory. The options are **FTP** and **SFTP**.
SFTP Public Key	Optional. This parameter specifies the public key of the FTP server and is valid only when **Dump Mode** is set to **SFTP**. You are advised to configure a public key. Otherwise, security risks may arise.

Step 4 Click **OK** to complete the settings.

----End

10.4 Alarm Management

10.4.1 Viewing and Manually Clearing an Alarm

Scenario

You can view and clear alarms on MRS Manager.

Generally, the system automatically clears an alarm when the fault is rectified. If the fault has been rectified and the alarm cannot be automatically cleared, you can manually clear the alarm.

You can view the latest 100,000 alarms (including uncleared, manually cleared, and automatically cleared alarms) on MRS Manager. If the number of cleared alarms exceeds 100,000 and is about to reach 110,000, the system automatically dumps the earliest 10,000 cleared alarms to **${BIGDATA_HOME}/OMSV100R001C00x8664/workspace/data** on the active management node. A directory is automatically generated when alarms are dumped for the first time.

NOTE

Set an automatic refresh interval or click ↻ for an immediate refresh.

The following refresh interval options are supported:

- Refresh every 30 seconds
- Refresh every 60 seconds
- Stop refreshing

Procedure

Step 1 On MRS Manager, click **Alarms** to view the alarm information in the alarm list.

- By default, the alarm list page displays the latest 10 alarms.
- By default, alarms are displayed in descending order by **Generated**. You can click **Alarm ID**, **Alarm Name**, **Severity**, **Generated**, **Location**, **Operation** to change the display mode.
- You can filter all alarms of the same severity in **Severity**, including cleared and uncleared alarms.
- You can click 🔴, 🟠, 🟡, or 🔵 to filter out **Critical**, **Major**, **Minor**, or **Warning** alarms.

Step 2 Click **Advanced Search**. In the displayed alarm search area, set search criteria and click **Search** to view the information about specified alarms. Click **Reset** to clear the search criteria.

NOTE

You can set the **Start Time** and **End Time** to specify the time range. You can search for alarms generated within the time range.

Handle the alarm by referring to **Alarm Reference**. If the alarms in some scenarios are generated due to other cloud services that MRS depends on, you need to contact maintenance personnel of the corresponding cloud services.

Step 3 If the alarm needs to be manually cleared after errors are rectified, click **Clear Alarm**.

NOTE

If multiple alarms have been handled, you can select one or more alarms to be cleared and click **Clear Alarm** to clear the alarms in batches. A maximum of 300 alarms can be cleared in each batch.

----End

10.4.2 Configuring an Alarm Threshold

Scenario

You can configure an alarm threshold to learn the metric health status. After **Send Alarm** is selected, the system sends an alarm message when the monitored data reaches the alarm threshold. You can view the alarm information in **Alarms**.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 In **Configuration**, click **Configure Alarm Threshold** under **Monitoring and Alarm**, select monitoring metrics as planned, and set their baselines.

Step 3 Click a metric, for example, **CPU Usage**, and click **Create Rule**.

Step 4 Set the monitoring metric rule parameters on the displayed configuration page.

Table 10-13 Monitoring metric rule parameters

Parameter	Value	Description
Rule Name	CPU_MAX (example value)	Specifies the rule name.
Reference Date	2014/11/06 (example)	Specifies the date on which the reference indicator history is generated.

Parameter	Value	Description
Threshold Type	● Max. value ● Min. value	Specifies the maximum or minimum value of a metric. If this parameter is set to **Max. Value**, the system generates an alarm when the actual value of the metric is greater than the threshold. If this parameter is set to **Min. Value**, the system generates an alarm when the actual value of the metric is smaller than the threshold.
Alarm Severity	● Critical ● Major ● Minor ● Suggestion	Alarm Severity
Time Range	From 00:00 to 23:59 (example)	Specifies the period in which the rule takes effect.
Threshold	80 (example)	Specifies the threshold of the rule monitoring metrics.
Date	● Workday ● Weekend ● Other	Specifies the type of date when the rule takes effect.
Add Date	11/06 (example)	This parameter is valid only when **Date** is set to **Other**. You can select multiple dates.

Step 5 Click **OK**. A message is displayed in the upper right corner of the page, indicating that the template is saved successfully.

Send alarm is selected by default. MRS Manager checks whether the value of each monitored metric reaches the threshold. If the number of consecutive check times is equal to the value of **Trigger Count**, and the threshold is not reached in these checks, the system sends an alarm. The value can be customized. **Check Period (s)** indicates the interval at which MRS Manager checks monitoring metrics.

Step 6 Locate the row that contains the newly added rule, and click **Apply** in the **Operation** column. A message is displayed in the upper right corner, indicating that the rule *xx* is successfully added. Click **Cancel** in the **Operation** column. A message is displayed in the upper right corner, indicating that the rule *xx* is successfully canceled.

----End

10.4.3 Configuring Syslog Northbound Interface Parameters

Scenario

You can configure the northbound interface so that alarms generated on MRS Manager can be reported to your monitoring O&M system using Syslog.

NOTICE

If the Syslog protocol is not encrypted, data may be stolen.

Prerequisites

The ECS corresponding to the server must be in the same VPC as the Master node of the MRS cluster, and the Master node can access the IP address and specified port of the server.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 In **Configuration**, click **Configure Syslog** under **Monitoring and Alarm**.

The **Syslog Service** is disabled by default. Click the switch to enable the Syslog service.

Step 3 Set the interconnection parameters listed in **Table 10-14**.

Table 10-14 Syslog parameters

Area	Parameter	Description
Syslog Protocol	Service IP Address	Specifies the IP address of the interconnection server.
	Server Port	Specifies the port number for interconnection.
	Protocol	Specifies the protocol type. The options are as follows: ● **TCP** ● **UDP**

Area	Parameter	Description
	Severity	Specifies the severity of the reported message. The options are as follows: ● **Informational** ● **Emergency** ● **Alert** ● **Critical** ● **Error** ● **Warning** ● **Notice** ● **Debug**
	Facility	Specifies the module where the log is generated.
	Identifier	Specifies the product ID. The default value is **MRS Manager**.
Report Message	Report Format	Specifies the message format of the alarm report. For details, see help information on the web page.
	Alarm Status	Specifies the type of the alarm to be reported. ● **Fault**: indicates that the Syslog alarm message is reported when MRS Manager generates an alarm. ● **Clear**: indicates that a Syslog alarm message is reported when an alarm on MRS Manager is cleared. ● **Event**: indicates that the Syslog alarm message is reported when MRS Manager generates an event.

Area	Parameter	Description
	Report Alarm Severity	Specifies the level of the alarm to be reported. The value can be **Suggestion**, **Minor**, **Major**, and **Critical**.
Uncleared Alarm Reporting	Periodic Uncleared Alarm Report	Specifies whether uncleared alarms are reported periodically. By default, the switch of **Periodic Uncleared Alarm Reporting** is disabled. You can click the switch to enable it.
	Report Interval (min)	Specifies the interval for periodically reporting uncleared alarms to the remote Syslog service. This parameter is valid only when **Periodic Uncleared Alarm Reporting** switch is enabled. The unit is minute. The default value is **15**. The value ranges from 5 minutes to one day (1,440 minutes).
Heartbeat Settings	Heartbeat Report	Specifies whether to periodically report Syslog heartbeat messages. By default, the switch of **Periodic Uncleared Alarm Reporting** is disabled. You can click the switch to enable it.
	Heartbeat Period (min)	Specifies the interval for periodically reporting heartbeat messages. This parameter is valid only when **Heartbeat Report** switch is enabled. The unit is minute. The default value is **15**. The value ranges from 1 to 60.

Area	Parameter	Description
	Heartbeat Packet	Specifies the content of the reported heartbeat message. This parameter is enabled when **Heartbeat Report** is enabled. The value can contain a maximum of 256 characters, including digits, letters, underscores (_), vertical bars (|), colons (:), spaces, commas (,), and periods (.).

NOTE

After the periodic heartbeat packet function is enabled, packets may be interrupted during automatic recovery of some cluster error tolerance (for example, active/standby management node switchover). In this case, wait for automatic recovery.

Step 4 Click **OK** to complete the settings.

----End

10.4.4 Configuring SNMP Northbound Interface Parameters

Scenario

You can configure the northbound interface so that alarms and monitoring metrics on MRS Manager can be integrated to the network management platform using SNMP.

Prerequisites

The ECS corresponding to the server must be in the same VPC as the Master node of the MRS cluster, and the Master node can access the IP address and specified port of the server.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 In **Configuration**, click **Configure SNMP** under **Monitoring and Alarm**.

The **SNMP Service** is disabled by default. Click the switch to enable the SNMP service.

Step 3 Set the interconnection parameters listed in **Table 10-15**.

Table 10-15 Syslog parameters

Parameter	Description
Version	Specifies the version of the SNMP, which can be: ● v2c: an earlier version with low security ● v3: the latest version of SNMP with higher security than SNMPv2c The SNMP v3 version is recommended.
Local Port	Specifies the local port. The default value is **20000**. The value ranges from **1025** to **65535**.
Read Community Name	Specifies the read-only community name. This parameter is valid only when **Version** is set to **v2c**.
Write Community Name	Specifies the write community name. This parameter is valid only when **Version** is set to **v2c**.
Security Username	Specifies the SNMP security username. This parameter is valid only when **Version** is set to **v3**.
Authentication Protocol	Specifies the authentication protocol. You are advised to set this parameter to set this parameter to **SHA**. This parameter is valid only when **Version** is set to **v3**.
Authentication Password	Specifies the authentication key. This parameter is valid only when **Version** is set to **v3**.
Confirm Password	Used to confirm the authentication key. This parameter is valid only when **Version** is set to **v3**.
Encryption Protocol	Specifies the encryption protocol. You are advised to set this parameter to **AES256**. This parameter is valid only when **Version** is set to **v3**.
Encryption Password	Specifies the encryption key. This parameter is valid only when **Version** is set to **v3**.
Confirm Password	Used to confirm the encryption key. This parameter is valid only when **Version** is set to **v3**.

NOTE

- The **Authentication Password** and **Encryption Password** must contain 8 to 16 characters, including at least three types of the following characters: uppercase letters, lowercase letters, digits, and special characters. The two passwords must be different. The two passwords cannot be the same as the security username or the reverse of the security username.
- For security purposes, periodically change the authentication password and encryption password when the SNMP protocol is used.
- If SNMPv3 is used, a security user will be locked after five consecutive authentication failures within 5 minutes. The user will be automatically unlocked 5 minutes later.

Step 4 Click **Create Trap Target** in the **Trap Target** area. In the displayed dialog box, set the following parameters:

- **Target Symbol** specifies the trap target ID, which is the ID of the NMS or host that receives traps. The value consists of 1 to 255 characters, including letters or digits.
- **Target IP Address** specifies the IP address of the target trap. IP addresses of class A, B, and C can be used to communicate with the IP address of the management plane of the management node.
- **Target Port** specifies the port receiving traps. The port number must be consistent with the peer end and ranges from 0 to 65535.
- **Trap Community Name** is valid only when **Version** is set to **v2c**.

Click **OK**. The **Create Trap Target** dialog box is closed.

Step 5 Click **OK** to complete the settings.

----End

10.5 Object Management

10.5.1 Managing Objects

MRS contains different types of basic objects as described in **Table 10-16**.

Table 10-16 MRS basic object overview

Object	Description	Example
Service	Function set that can complete specific business.	KrbServer service and LdapServer service
Service instance	Specific instance of a service, usually called service.	KrbServer service
Service role	Function entity that forms a complete service, usually called role.	KrbServer is composed of the KerberosAdmin role and KerberosServer role.
Role instance	Specific instance of a service role running on a host.	KerberosAdmin that is running on Host2 and KerberosServer that is running on Host3
Host	An ECS running Linux OS.	Host1 to Host5
Rack	Physical entity that contains multiple hosts connecting to the same switch.	Rack1 contains Host1 to Host5.
Cluster	Logical entity that consists of multiple hosts and provides various services.	Cluster names **Cluster1** consists of five hosts (Host1 to Host5) and provides services such as KrbServer and LdapServer.

10.5.2 Viewing Configurations

On MRS Manager, users can view the configurations of services (including roles) and role instances.

Procedure

- Query service configurations.
  a. On MRS Manager page, click **Services**.
  b. Select the target service from the service list.
  c. Click **Service Configuration**.
  d. Set **Type** to **All**. All configuration parameters of the service are displayed in the navigation tree. The root nodes from top down in the navigation tree represent the service names and role names.
  e. In the navigation tree, select a specified parameter and change its value. You can also enter the parameter name in the **Search** box to search for the parameter and view the result.

     The parameters under the service nodes and role nodes are service configuration parameters and role configuration parameters respectively.
  f. In the **Non-default** parameter, select **Non-default**. The parameters whose values are not default values will be displayed.
- Query role instance configurations.
  a. On MRS Manager page, click **Services**.
  b. Select the target service from the service list.
  c. Click the **Instances** tab.
  d. Click the target role instance from the role instance list.
  e. Click **Instance Configuration**.
  f. Set **Type** to **All**. The navigation tree of all configuration parameters of the role instance is displayed.
  g. In the navigation tree, select a specified parameter and change its value. You can also enter the parameter name in the **Search** box to search for the parameter and view the result.
  h. In the **Non-default** parameter, select **Non-default**. The parameters whose values are not default values will be displayed.

10.5.3 Managing Services

You can perform the following operations on MRS Manager:

- Start the service in the **Stopped**, **Stop Failed**, or **Start Failed** state to use the service.
- Stop the services or stop abnormal services.
- Restart abnormal services or configure expired services to restore or enable the services.

Procedure

Step 1 On MRS Manager page, click **Services**.

Step 2 Locate the row that contains the target service, **Start**, **Stop**, or **Restart** to start, stop, or restart the service.

Services are interrelated. If a service is started, stopped, and restarted, services dependent on it will be affected.

The services will be affected in the following ways:

- If a service is to be started, the lower-layer services dependent on it must be started first.
- If a service is stopped, the upper-layer services dependent on it are unavailable.
- If a service is restarted, the running upper-layer services dependent on it must be restarted.

----End

10.5.4 Configuring Service Parameters

On MRS Manager, you can view and modify the default service configurations based on site requirements and export or import the configurations.

Impact on the System

- You need to download and update the client configuration files after configuring HBase, HDFS, Hive, Spark, Yarn, and MapReduce service properties.
- The parameters of DBService cannot be modified when only one DBService role instance exists in the cluster.

Procedure

- Modify a service.
  a. Click **Services**.
  b. Select the target service from the service list.
  c. Click **Service Configuration**.
  d. Set **Type** to **All**. All configuration parameters of the service are displayed in the navigation tree. The root nodes from top down in the navigation tree represent the service names and role names.
  e. In the navigation tree, select a specified parameter and change its value. You can also enter the parameter name in the **Search** box to search for the parameter and view the result.

     If you want to cancel the modification of a parameter value, click ↻ to restore it.

NOTE

You can also use host groups to change role instance configurations in batches. Select a role name from the **Role** drop-down list and choose **< Select Host >** in the **Host** drop-down list. Enter a name in the **Host Group Name** text box, select the hosts to be modified from the **Host** list, add them to the **Selected hosts** area, and click **OK**. The added host group can be selected from **Host** and is only valid on the current page. The page cannot be saved after being refreshed.

f. Click **Save Configuration** and select **Restart the affected services or instances**. Click **OK** to restart the services.

   After **Operation successful.** is displayed, click **Finish**. The service is started successfully.

NOTE

To update the queue configuration of the Yarn service without restarting service, choose **More** > **Refresh Queue** to update the queue for the configuration to take effect.

- Export service configuration parameters.
  a. Click **Services**.
  b. Select a service.
  c. Click **Service Configuration**.
  d. Click **Export Service Configuration**. Select a path for saving the configuration files.
- Import service configuration parameters.
  a. Click **Services**.
  b. Select a service.
  c. Click **Service Configuration**.
  d. Click **Import Service Configuration**.
  e. Select the target configuration file.
  f. Click **Save Configuration** and select **Restart the affected services or instances**. Click **OK**.

     After **Operation successful.** is displayed, click **Finish**. The service is started successfully.

10.5.5 Configuring Customized Service Parameters

Each component of MRS supports all open-source parameters. You can modify some parameters for key application scenarios on MRS Manager. Some component clients may not include all parameters with open-source features. For component parameters that cannot be directly modified on Manager, users can add new parameters for components by using the configuration customization function on Manager. Newly added parameters are saved in component configuration files and take effect after restart.

Impact on the System

- After the service attributes are configured, the service needs to be restarted and cannot be accessed.

- You need to download and update the client configuration files after configuring HBase, HDFS, Hive, Spark, Yarn, and MapReduce service properties.

Prerequisites

You have understood the meanings of parameters to be added, configuration files that have taken effect, and the impact on components.

Procedure

Step 1 On MRS Manager, click **Services**.

Step 2 Select the target service from the service list.

Step 3 Click **Service Configuration**.

Step 4 Set **Type** to **All**.

Step 5 In the navigation tree, select **Customization**. The customized parameters of the current component are displayed on Manager.

The configuration files that save the newly added customized parameters are displayed in the **Parameter File** column. Different configuration files may have same open source parameters. After the parameters in different files are set to different values, whether the configuration takes effect depends on the loading sequence of the configuration files by components. You can customize parameters for services and roles as required. Adding customized parameters for a single role instance is not supported.

Step 6 Based on the configuration files and parameter functions, locate the row where a specified parameter resides, enter the parameter name supported by the component in the **Name** column and enter the parameter value in the **Value** column.

- You can click ⊕ or ⊗ to add or delete a user-defined parameter. You can delete a customized parameter only after you click ⊕ for the first time.
- If you want to cancel the modification of a parameter value, click ↻ to restore it.

Step 7 Click **Save Configuration** and select **Restart the affected services or instances**. Click **OK** to restart the services.

After **Operation successful.** is displayed, click **Finish**. The service is started successfully.

----End

Task Example

Configuring Customized Hive Parameters

Hive depends on HDFS. By default, Hive accesses the HDFS client. The configuration parameters to take effect are controlled by HDFS in a unified manner. For example, the HDFS parameter **ipc.client.rpc.timeout** affects the RPC timeout period for all clients to connect to the HDFS server. If you need to modify the timeout period for Hive to connect to HDFS, you can use the configuration customization function. After this parameter is added to the **core-site.xml** file of Hive, this parameter can be identified by the Hive service and its configuration overwrites the parameter configuration in HDFS.

Step 1 On MRS Manager, choose **Services** > **Hive** > **Service Configuration**.

Step 2 Set **Type** to **All**.

Step 3 In the navigation tree on the left, select **Customization** for the Hive service. The system displays the customized service parameters supported by Hive.

Step 4 In **core-site.xml**, locate the row that contains the **core.site.customized.configs** parameter, enter **ipc.client.rpc.timeout** in the **Name** column, and enter a new value in the **Value** column, for example, **150000**. The unit is millisecond.

Step 5 Click **Save Configuration** and select **Restart the affected services or instances**. Click **OK** to restart the service.

After **Operation successful.** is displayed, click **Finish**. The service is started successfully.

----End

10.5.6 Synchronizing Service Configurations

Scenario

If **Configuration Status** of a service is **Expired** or **Failed**, synchronize configurations for the cluster or service to restore its configuration status. If all services in the cluster are in the **Failed** state, synchronize the cluster configuration with the background configuration.

Impact on the System

After synchronizing service configurations, you need to restart the services whose configurations have expired. These services are unavailable during restart.

Procedure

Step 1 On MRS Manager page, click **Services**.

Step 2 Select the target service from the service list.

Step 3 In the upper part of the service status and metric information, choose **More** > **Synchronize Configuration**.

Step 4 In the displayed dialog box, select **Restart services and instances whose configuration have expired.** and click **OK** to restart the service whose configuration has expired.

When **Operation successful** is displayed, click **Finish**. The service is started successfully.

----End

10.5.7 Managing Role Instances

Scenario

You can start a role instance that is in the **Stopped**, **Failed to stop** or **Failed to start** status, stop an unused or abnormal role instance or restart an abnormal role instance to recover its functions.

Procedure

Step 1 On MRS Manager page, click **Services**.

Step 2 Select the target service from the service list.

Step 3 Click the **Instances** tab.

Step 4 Select the check box on the left of the target role instance.

Step 5 Choose **More** > **Start Instance**, **Stop Instance**, or **Restart Instance** accordingly.

----End

10.5.8 Configuring Role Instance Parameters

Scenario

You can view and modify default role instance configurations on MRS Manager based on site requirements. The configurations can be imported and exported.

Impact on the System

You need to download and update the client configuration files after configuring HBase, HDFS, Hive, Spark, Yarn, and MapReduce service properties.

Procedure

- Modifying role instance configurations
  a. Click **Services**.
  b. Select the target service from the service list.
  c. Click the **Instances** tab.
  d. Click the target role instance from the role instance list.
  e. Click **Instance Configuration**.
  f. Set **Type** to **All**. The navigation tree of all configuration parameters of the role instance is displayed.
  g. In the navigation tree, select a specified parameter and change its value. You can also enter the parameter name in the **Search** box to search for the parameter and view the result.

     If you want to cancel the modification of a parameter value, click ↻ to restore it.
  h. Click **Save Configuration**, select **Restart the role instance**, and click **OK** to restart the role instance.

After **Operation successful.** is displayed, click **Finish**. The role instance is started successfully.

- Exporting Configuration Parameters of a Role Instance
  a. Click **Services**.
  b. Select a service.
  c. Select a role instance or click the **Instances** tab.
  d. Select a role instance on a specified host.
  e. Click **Instance Configuration**.
  f. Click **Export Instance Configuration** to export the configuration data of a specified role instance, and choose a path for saving the configuration file.
- Import configuration data of a role instance.
  a. Click **Services**.
  b. Select a service.
  c. Select a role instance or click the **Instances** tab.
  d. Select a role instance on a specified host.
  e. Click **Instance Configuration**.
  f. Click **Import Instance Configuration** to import the configuration data of the specified role instance.
  g. Click **Save Configuration** and select **Restart the role instance**. Click **OK**.

     After **Operation successful.** is displayed, click **Finish**. The role instance is started successfully.

10.5.9 Synchronizing Role Instance Configuration

Scenario

When **Configuration Status** of a role instance is **Expired** or **Failed**, you can synchronize the configuration data of the role instance with the background configuration.

Impact on the System

After synchronizing a role instance configuration, you need to restart the role instance whose configuration has expired. The role instance is unavailable during restart.

Procedure

Step 1 On MRS Manager, click **Services** and select a service name.

Step 2 Click the **Instances** tab.

Step 3 Click the target role instance from the role instance list.

Step 4 Choose **More** > **Synchronize Configuration** above the role instance status and indicator information.

Step 5 In the displayed dialog box, select **Restart services and instances whose configuration have expired**, and click **OK** to restart the role instance.

After **Operation successful** is displayed, click **Finish**. The role instance is started successfully.

----End

10.5.10 Decommissioning and Recommissioning a Role Instance

Scenario

If a Core or Task node is faulty, the cluster status may be displayed as **Abnormal**. In an MRS cluster, data can be stored on different Core nodes. Users can decommission the specified role instance on MRS Manager to stop the role instance from providing services. After fault rectification, you can recommission the role instance.

The following role instances can be decommissioned and recommissioned.

- DataNode role instance on HDFS
- NodeManager role instance on Yarn
- RegionServer role instance on HBase
- Broker role instance on Kafka

Restrictions:

- If the number of the DataNodes is less than or equal to that of HDFS copies, decommissioning cannot be performed. If the number of HDFS copies is three and the number of DataNodes is less than four in the system, decommissioning cannot be performed. In this case, an error will be reported and the decommissioning will be stopped 30 minutes after the decommissioning attempt is performed on Manager.
- If the number of Kafka Broker instances is less than or equal to that of copies, decommissioning cannot be performed. For example, if the number of Kafka copies is two and the number of nodes is less than three in the system, decommissioning cannot be performed. Instance decommissioning will fail on Manager and exit.
- If a role instance is out of service, you must recommission the instance to start it before using it again.

Procedure

Step 1 On MRS Manager page, click **Services**.

Step 2 Click a service in the service list.

Step 3 Click the **Instances** tab.

Step 4 Select an instance.

Step 5 Choose **More** > **Decommission** or **Recommission** to perform the corresponding operation.

NOTE

During the instance decommissioning, if the service corresponding to the instance is restarted in the cluster using another browser, MRS Manager displays a message indicating that the instance decommissioning is stopped, but the **Operating Status** of the instance is displayed as **Started**. In this case, the instance has been decommissioned on the background. You need to decommission the instance again to synchronize the operating status.

----End

10.5.11 Managing a Host

Scenario

When a host is abnormal or faulty, you need to stop all roles of the host on MRS Manager to check the host. After the host fault is rectified, start all roles running on the host to recover host services.

Procedure

Step 1 Click **Hosts**.

Step 2 Select the check box of the target host.

Step 3 Choose **More** > **Start All Roles** or **Stop All Roles** accordingly.

----End

10.5.12 Isolating a Host

Scenario

If a host is found to be abnormal or faulty, affecting cluster performance or preventing services from being provided, you can temporarily exclude that host from the available nodes in the cluster. In this way, the client can access other available nodes. In scenarios where patches are to be installed in a cluster, you can also exclude a specified node from patch installation.

Users can isolate a host manually on MRS Manager based on the actual service requirements or O&M plan. Only non-management nodes can be isolated.

Impact on the System

- After a host is isolated, all role instances on the host will be stopped. You cannot start, stop, or configure the host and any instances on the host.
- After a host is isolated, statistics about the monitoring status and indicator data of the host hardware and instances on the host cannot be collected or displayed.

Procedure

Step 1 On MRS Manager, click **Hosts**.

Step 2 Select the check box of the host to be isolated.

Step 3 Choose **More** > **Isolate Host**,

Step 4 and click **OK** in the displayed dialog box.

After **Operation successful.** is displayed, click **Finish**. The host is isolated successfully, and the value of **Operating Status** becomes **Isolated**.

NOTE

For isolated hosts, you can cancel the isolation and add them to the cluster again. For details, see **Canceling Host Isolation**.

----End

10.5.13 Canceling Host Isolation

Scenario

After the exception or fault of a host is handled, you must cancel the isolation of the host for proper usage.

Users can cancel the isolation of a host on MRS Manager.

Prerequisites

- The host is in the **Isolated** state.
- The exception or fault of the host has been rectified.

Procedure

Step 1 On MRS Manager, click **Hosts**.

Step 2 Select the check box of the host to be de-isolated.

Step 3 Choose **More** > **Cancel Host Isolation**,

Step 4 and click **OK** in the displayed dialog box.

After **Operation successful.** is displayed, click **Finish**. The host is de-isolated successfully, and the value of **Operating Status** becomes **Normal**.

Step 5 Click the name of the de-isolated host to show its status, and click **Start All Roles**.

----End

10.5.14 Starting or Stopping a Cluster

Scenario

A cluster is a collection of service components. You can start or stop all services in a cluster.

Procedure

Step 1 On MRS Manager page, click **Services**.

Step 2 In the upper part of the service list, choose **More** > **Start Cluster** or **Stop Cluster** accordingly.

----End

10.5.15 Synchronizing Cluster Configurations

Scenario

If **Configuration Status** of all services or some services is **Expired** or **Failed**, synchronize configuration for the cluster or service to restore its configuration status.

- If all services in the cluster are in the **Failed** state, synchronize the cluster configuration with the background configuration.
- If all services in the cluster are in the **Failed** state, synchronize the service configuration with the background configuration.

Impact on the System

After synchronizing cluster configurations, you need to restart the services whose configurations have expired. These services are unavailable during restart.

Procedure

Step 1 On MRS Manager page, click **Services**.

Step 2 In the upper part of the service list, choose **More** > **Synchronize Configuration**.

Step 3 In the displayed dialog box, select **Restart services and instances whose configuration have expired**, and click **OK** to restart the service whose configuration has expired.

When **Operation successful.** is displayed, click **Finish**. The service is started successfully.

----End

10.5.16 Exporting Configuration Data of a Cluster

Scenario

You can export all configuration data of a cluster on MRS Manager to meet site requirements. The exported configuration data is used to rapidly update service configuration.

Procedure

Step 1 On MRS Manager page, click **Services**.

Step 2 Choose **More** > **Export Cluster Configuration**.

The exported file is used to update service configurations. For details, see **Import service configuration parameters** in Configuring Service Parameters.

----End

10.6 Log Management

10.6.1 About Logs

Log Description

MRS cluster logs are stored in the **/var/log/Bigdata** directory. The following table lists the log types.

Table 10-17 Log types

Type	Description
Installation log	Installation logs record information about FusionInsight Manager, cluster, and service installation to help users locate installation errors.
Run logs	Run logs record the running track information, debugging information, status changes, potential problems, and error information generated during the running of services.
Audit logs	Audit logs record information about users' activities and operation instructions, which can be used to locate fault causes in security events and determine who are responsible for these faults.

The following table lists the MRS log directories.

Table 10-18 Log directories

File Directory	Log Content
/var/log/Bigdata/audit	Component audit log.
/var/log/Bigdata/controller	Log collecting script log. Controller process log. Controller monitoring log.
/var/log/Bigdata/dbservice	DBService log.
/var/log/Bigdata/flume	Flume log.
/var/log/Bigdata/hbase	HBase log.
/var/log/Bigdata/hdfs	HDFS log.
/var/log/Bigdata/hive	Hive log.

File Directory	Log Content
/var/log/Bigdata/httpd	HTTPD log.
/var/log/Bigdata/hue	Hue log.
/var/log/Bigdata/kerberos	Kerberos log.
/var/log/Bigdata/ldapclient	LDAP client log.
/var/log/Bigdata/ldapserver	LDAP server log.
/var/log/Bigdata/loader	Loader log.
/var/log/Bigdata/logman	logman script log management log.
/var/log/Bigdata/mapreduce	MapReduce log.
/var/log/Bigdata/nodeagent	NodeAgent log.
/var/log/Bigdata/okerberos	OMS Kerberos log.
/var/log/Bigdata/oldapserver	OMS LDAP log.
/var/log/Bigdata/omm	**oms**: complex event processing log, alarm service log, HA log, authentication and authorization management log, and monitoring service run log of the omm server. **oma**: installation log and run log of the omm agent. **core**: dump log generated when the omm agent and the HA process are suspended.
/var/log/Bigdata/spark	Spark log.
/var/log/Bigdata/sudo	Log generated when the **sudo** command is executed by user **omm**.
/var/log/Bigdata/timestamp	Time synchronization management log.
/var/log/Bigdata/tomcat	Tomcat log.
/var/log/Bigdata/yarn	Yarn log.
/var/log/Bigdata/zookeeper	ZooKeeper log.
/var/log/Bigdata/kafka	Kafka log.
/var/log/Bigdata/storm	Storm log.
/var/log/Bigdata/patch	Patch log.

Run logs

Table 10-19 describes the running information recorded in run logs.

Table 10-19 Running information

Run Log	Description
Installation preparation log	Records information about preparations for the installation, such as the detection, configuration, and feedback operation information.
Process startup log	Records information about the commands executed during the process startup.
Process startup exception log	Records information about exceptions during process startup, such as dependent service errors and insufficient resources.
Process run log	Records information about the process running track information and debugging information, such as function entries and exits as well as cross-module interface messages.
Process running exception log	Records errors that cause process running errors, for example, the empty input objects or encoding or decoding failure.
Process running environment log	Records information about the process running environment, such as resource status and environment variables.
Script logs	Records information about the script execution process.
Resource reclamation log	Records information about the resource reclaiming process.
Uninstallation clearing logs	Records information about operations performed during service uninstallation, such as directory deletion and execution time

Audit logs

Audit information recorded in audit logs includes FusionInsight Manager audit information and component audit information.

Table 10-20 Audit information of FusionInsight Manager

Audit Log	Operation Type	Operation
Manager audit log	User management	Creating a user Modifying a user Deleting a user Creating a user group Modifying a user group Deleting a user group Adding a role Modifying a role Deleting a role Changing a password policy Changing a password Resetting a password User login User logout Unlocking the screen Downloading the authentication credential Unauthorized operation Unlocking a user account Locking a user account Locking the screen Exporting user information Exporting a user group Exporting a role

Audit Log	Operation Type	Operation
	Tenant management	Saving the static configuration Adding a tenant Deleting a tenant Associating a service with a tenant Deleting a service from a tenant Configuring resources Creating resources Deleting resources Adding a resource pool Modifying a resource pool Deleting a resource pool Restoring tenant data

Audit Log	Operation Type	Operation
	Cluster management	Starting a cluster Stopping a cluster Saving configurations Synchronizing cluster configurations Customizing cluster monitoring indicators Saving monitoring thresholds Downloading a client configuration file Configuring the northbound API Configuring the northbound SNMP API Creating a threshold template Deleting a threshold template Applying a threshold template Saving cluster monitoring configuration data Exporting configuration data Importing cluster configuration data Exporting an installation template Modifying a threshold template Canceling the application of a threshold template Masking alarms Sending an alarm Changing the OMS database password Changing the component database password Starting the health check of a cluster

Audit Log	Operation Type	Operation
		Updating the health check configuration Exporting cluster health check results Importing a certificate file Deleting historical health check reports Exporting historical health check reports Customizing report monitoring indicators Exporting report monitoring data Customizing monitoring indicators for static resource pools Exporting monitoring data of a static resource pool
	Service management	Starting a service Stopping a service Synchronizing service configurations Refreshing a service queue Customizing service monitoring indicators Restarting a service Exporting service monitoring data Importing service configuration data Starting the health check of a service Exporting service health check results Configuring the service Uploading a configuration file Downloading a configuration file

Audit Log	Operation Type	Operation
	Instance management	Synchronizing instance configurations Commissioning an instance Decommissioning an instance Starting an instance Stopping an instance Customizing instance monitoring indicators Restarting an instance Exporting instance monitoring data Importing instance configuration data
	Host management	Setting a node rack Starting all roles Stopping all roles Isolating a host Canceling host isolation Customizing host monitoring indicators Exporting host monitoring data Starting the health check of a host Exporting the health check result of a host

Audit Log	Operation Type	Operation
	Maintenance management	Exporting alarms Clearing alarms Exporting events Clearing alarms in batches Clearing alarm through SNMP Adding a trap target through SNMP Deleting a trap target through SNMP Checking alarms through SNMP Synchronizing alarms through SNMP Modifying audit dump configurations Exporting audit logs Collecting log files Downloading log files Uploading a file Deleting an uploaded file Creating a backup task Executing a backup task Stopping a backup task Deleting a backup task Modifying a backup task Locking a backup task Unlocking a backup task Creating a restoration task Executing a backup restoration task Stopping a restoration task Retrying a restoration task Deleting a restoration task

Table 10-21 Component audit information

Audit Log	Operation Type	Operation
DBService audit log	Maintenance management	Performing backup restoration operations
HBase audit log	Data definition language (DDL) statement	Creating a table Deleting a table Modifying a table Adding a column family Modifying a column family Deleting a column family Enabling a table Disabling a table Modify the user information Changing a password User login
	Data manipulation language (DML) statement	Putting data (to the **hbase:meta**, **_ctmeta_**, and **hbase:acl** tables) Deleting data (from the **hbase:meta**, **_ctmeta_**, and **hbase:acl** tables) Checking and putting data (to the **hbase:meta**, **_ctmeta_**, and **hbase:acl** tables) Checking and deleting data (from the **hbase:meta**, **_ctmeta_**, and **hbase:acl** tables)
	Permission control	Assigning permissions to a user Canceling permission assigning

Audit Log	Operation Type	Operation
Hive audit logs	Metadata operation	Defining metadata, such as creating databases and tables Deleting metadata, such as deleting databases and tables Modifying metadata, such as adding columns and renaming tables Importing and exporting metadata
	Data maintenance	Loading data to a table Inserting data into a table
	Permissions management	Creating or deleting roles Granting/Reclaiming roles Granting/Reclaiming permissions
HDFS audit log	Permissions management	Managing permissions on files or folders Managing permissions on owner information files or folders
	File operation	Creating a folder Creating a file Opening a file Appending file content Changing a file name Deleting a file or folder Setting time property of a file Setting the number of file copies Merging files Checking the file system File links

Audit Log	Operation Type	Operation
MapReduce audit log	Application running	Starting a Container request Stopping a Container request After Container request is completed, the status of the request is displayed as succeeded. After Container request is completed, the status of the request is displayed as failed. After Container request is completed, the status of the request is displayed as suspended. Submitting a task Ending a task
LdapServer audit log	Maintenance management	Adding an operating system user Adding a user group Adding a user to user group Deleting a user Deleting a group
KrbServer audit log	Maintenance management	Changing the password of a Kerberos account Adding a Kerberos account Deleting a Kerberos account Authenticating a user
Loader audit log	Security management	User login
	Metadata management	Querying connector information Querying a framework Querying step information

Audit Log	Operation Type	Operation
	Managing data source connections	Querying a data source connection Adding a data source connection Updating a data source connection Deleting a data source connection Activating a data source connection Disabling a data source connection
	Job management	Querying a job Creating a Job Updating a Job Deleting a job Activating a job Disabling a job Querying all execution records of a job Querying the latest execution record of a job Submitting a job Stopping a job
Hue audit log	Service startup	Starting Hue
	User operation	User login User logout
	Task operation	Creating a job Modifying a job Deleting a job Submitting a task Saving a task Updating the status of a task
ZooKeeper audit log	Permissions management	Setting the access permission to Znode
	Znode operation	Creating a Znode Deleting a Znode Configuring Znode data

Audit Log	Operation Type	Operation
Storm audit log	Nimbus	Submitting a topology Stopping a topology Reallocating a topology Deactivating a topology Activating a topology
	UI	Stopping a topology Reallocating a topology Deactivating a topology Activating a topology

MRS audit logs are stored in the database. You can view and export audit logs on the **Audit** page.

The following table lists the directories to store component audit logs. Audit log files of some components are stored in **/var/log/Bigdata/audit**, such as HDFS, HBase, MapReduce, Hive, Hue, Yarn, Storm, and ZooKeeper. The component audit logs are automatically compressed and backed up to **/var/log/Bigdata/audit/bk** at 03: 00 every day. A maximum of latest 90 compressed backup files are retained, and the backup time cannot be changed.

Audit log files of other components are stored in the component log directory.

Table 10-22 Directory for storing component audit logs

Component	Audit Log Directory
DBService	/var/log/Bigdata/audit/dbservice/dbservice_audit.log
HDFS	/var/log/Bigdata/audit/hdfs/nn/hdfs-audit-namenode.log /var/log/Bigdata/audit/hdfs/dn/hdfs-audit-datanode.log /var/log/Bigdata/audit/hdfs/jn/hdfs-audit-journalnode.log /var/log/Bigdata/audit/hdfs/zkfc/hdfs-audit-zkfc.log /var/log/Bigdata/audit/hdfs/httpfs/hdfs-audit-httpfs.log /var/log/Bigdata/audit/hdfs/router/hdfs-audit-router.log
MapReduce	/var/log/Bigdata/audit/mapreduce/jobhistory/mapred-audit-jobhistory.log

Component	Audit Log Directory
Hive	/var/log/Bigdata/audit/hive/hiveserver/hive-audit.log /var/log/Bigdata/audit/hive/metastore/metastore-audit.log /var/log/Bigdata/audit/hive/webhcat/webhcat-audit.log
Loader	/var/log/Bigdata/loader/audit/default.audit
Hue	/var/log/Bigdata/audit/hue/hue-audits.log
ZooKeeper	/var/log/Bigdata/audit/zookeeper/quorumpeer/zk-audit-quorumpeer.log
Spark	/var/log/Bigdata/audit/spark/jdbcserver/jdbcserver-audit.log /var/log/Bigdata/audit/spark/jobhistory/jobhistory-audit.log
Yarn	/var/log/Bigdata/audit/yarn/rm/yarn-audit-resourcemanager.log /var/log/Bigdata/audit/yarn/nm/yarn-audit-nodemanager.log
Storm	/var/log/Bigdata/audit/storm/nimbus/audit.log /var/log/Bigdata/audit/storm/ui/audit.log

10.6.2 Manager Log List

Log Description

Log path: The default storage path of Manager log files is **/var/log/Bigdata/***Manager component*.

- ControllerService: **/var/log/Bigdata/controller/** (operation & maintenance system (OMS) installation and run logs)
- Httpd: **/var/log/Bigdata/httpd** (httpd installation and run logs)
- logman: **/var/log/Bigdata/logman** (log packaging tool logs)
- NodeAgent: **/var/log/Bigdata/nodeagent** (NodeAgent installation and run logs)
- okerberos: **/var/log/Bigdata/okerberos** (okerberos installation and run logs)
- oldapserver: **/var/log/Bigdata/oldapserver** (oldapserver installation and run logs)
- MetricAgent: **/var/log/Bigdata/metric_agent** (MetricAgent run log)
- omm: **/var/log/Bigdata/omm** (omm installation and run logs)
- timestamp: **/var/log/Bigdata/timestamp** (NodeAgent startup time logs)
- tomcat: **/var/log/Bigdata/tomcat** (Web process logs)

- Patch: **/var/log/Bigdata/patch** (patch installation log)
- Sudo: **/var/log/Bigdata/sudo** (sudo script execution log)
- OS: **/var/log/***message file* (OS system log)
- OS Performance: **/var/log/osperf** (OS performance statistics log)
- OS Statistics: **/var/log/osinfo/statistics** (OS parameter configuration log)

Log archiving rule:

The automatic compression and archiving function is enabled for Manager logs. By default, when the size of a log file exceeds 10 MB, the log file is automatically compressed. The naming rule of a compressed log file is as follows: <*Original log name*>-<*yyyy-mm-dd_hh-mm-ss*>.[*ID*].**log.zip** A maximum of 20 latest compressed files are reserved.

Table 10-23 Manager logs

Type	Log File Name	Description
Controller run log	controller.log	Log that records component installation, upgrade, patch installation, configuration, monitoring, alarms, and routine O&M operations
	controller_client.log	Run log of the Representational State Transfer (REST) API
	acs.log	ACS run log file
	acs_spnego.log	spnego user log in ACS
	aos.log	AOS run log
	plugin.log	AOS plug-in log
	backupplugin.log	Log that records the backup and restoration operations
	controller_config.log	Configuration run log
	controller_nodesetup.log	Controller loading task log
	controller_root.log	System log of the Controller process
	controller_trace.log	Log that records the remote procedure call (RPC) communication between Controller and NodeAgent

Type	Log File Name	Description
	controller_monitor.log	Monitoring log
	controller_fsm.log	State machine log
	controller_alarm.log	Controller alarm log
	controller_backup.log	Controller backup and recovery log
	install.log, distributeAdapterFiles.log, install_os_optimization.log	OMS installation log
	oms_ctl.log	OMS startup and stop log
	installntp.log	NTP installation log
	modify_manager_param.log	Manager parameter modification log
	backup.log	OMS backup script run log
	supressionAlarm.log	Alarm script run log
	om.log	OM certificate generation log
	backupplugin_ctl.log	Startup log of the backup and restoration plug-in process
	getLogs.log	Run log of the collection log script
	backupAuditLogs.log	Run log of the audit log backup script
	certStatus.log	Log that records regular certificate checks
	distribute.log	Certificate distribution log
	ficertgenetrate.log	Certificate replacement logs, including logs of level-2 certificates, CAS certificates, and httpd certificates
	genPwFile.log	Log that records the generation of certificate password files

Type	Log File Name	Description
	modifyproxyconf.log	Log that records the modification of the HTTPD proxy configuration
	importTar.log	Log that records the process of importing certificates into the trust library
Httpd	install.log	Httpd installation log
	access_log, error_log	Httpd run log
logman	logman.log	Log packaging tool log
NodeAgent	install.log, install_os_optimization.log	NodeAgent installation log
	installntp.log	NTP installation log
	start_ntp.log	NTP startup log
	ntpChecker.log	NTP check log
	ntpMonitor.log	NTP monitoring log
	heartbeat_trace.log	Log that records heartbeats between NodeAgent and Controller
	alarm.log	Alarm log
	monitor.log	Monitoring log
	nodeagent_ctl.log, start-agent.log	NodeAgent startup log
	agent.log	NodeAgent run log
	cert.log	Certificate log
	agentplugin.log	Agent plug-in running status monitoring log
	omaplugin.log	OMA plug-in run log
	diskhealth.log	Disk health check log
	supressionAlarm.log	Alarm script run log
	updateHostFile.log	Host list update log
	collectLog.log	Run log of the node log collection script

Type	Log File Name	Description
	host_metric_collect.log	Host index collection run log
	checkfileconfig.log	Run log file of file permission check
	entropycheck.log	Entropy check run log
	timer.log	Log of periodic node scheduling
	pluginmonitor.log	Component monitoring plug-in log
	agent_alarm_py.log	Log that records alarms upon insufficient NodeAgent file permission
okerberos	addRealm.log, modifyKerberosRealm.log	Domain handover log
	checkservice_detail.log	Okerberos health check log
	genKeytab.log	keytab generation log
	KerberosAdmin_genConfigDetail.log	Run log that records the generation of kadmin.conf when starting the kadmin process
	KerberosServer_genConfigDetail.log	Run log that records the generation of krb5kdc.conf when starting the krb5kdc process
	oms-kadmind.log	Run log of the kadmin process
	oms_kerberos_install.log, postinstall_detail.log	Okerberos installation log
	oms-krb5kdc.log	Run log of the krbkdc process
	start_detail.log	Okerberos startup log
	realmDataConfigProcess.log	Log rollback for domain handover failure
	stop_detail.log	Okerberos stop log
oldapserver	ldapserver_backup.log	Oldapserver backup log

Type	Log File Name	Description
	ldapserver_chk_service.log	Oldapserver health check log
	ldapserver_install.log	Oldapserver installation log
	ldapserver_start.log	Oldapserver startup log
	ldapserver_status.log	Log that records the status of the Oldapserver process
	ldapserver_stop.log	Oldapserver stop log
	ldapserver_wrap.log	Oldapserver service management log
	ldapserver_uninstall.log	Oldapserver uninstallation log
	restart_service.log	Oldapserver restart log
	ldapserver_unlockUser.log	Log that records information about unlocking LDAP users and managing accounts
omm	omsconfig.log	OMS configuration log
	check_oms_heartbeat.log	OMS heartbeat log
	monitor.log	OMS monitoring log
	ha_monitor.log	HA_Monitor operation log
	ha.log	HA operation log
	fms.log	Alarm log
	fms_ha.log	HA alarm monitoring log
	fms_script.log	Alarm control log
	config.log	Alarm configuration log
	iam.log	IAM log
	iam_script.log	IAM control log
	iam_ha.log	IAM HA monitoring log
	config.log	IAM configuration log
	operatelog.log	IAM operation log

Type	Log File Name	Description
	heartbeatcheck_ha.log	OMS heartbeat HA monitoring log
	install_oms.log	OMS installation log
	pms_ha.log	HA monitoring log
	pms_script.log	Monitoring control log
	config.log	Monitoring configuration log
	plugin.log	Monitoring plug-in run log
	pms.log	Monitoring log
	ha.log	HA run log
	cep_ha.log	CEP HA monitoring log
	cep_script.log	CEP control log
	cep.log	CEP log
	config.log	CEP configuration log
	omm_gaussdba.log	GaussDB HA monitoring log
	gaussdb-<SERIAL>.log	GaussDB run log
	gs_ctl-<DATE>.log	GaussDB control log archive log
	gs_ctl-current.log	GaussDB control log
	gs_guc-current.log	GaussDB operation log
	encrypt.log	Omm encryption log
	omm_agent_ctl.log	OMA control log
	oma_monitor.log	OMA monitoring log
	install_oma.log	OMA installation log
	config_oma.log	OMA configuration log
	omm_agent.log	OMA run log
	acs.log	ACS resource log
	aos.log	AOS resource log
	controller.log	Controller resource log

Type	Log File Name	Description
	feed_watchdog.log	feed_watchdog resource log
	floatip.log	Floating IP address resource log
	ha_ntp.log	NTP resource log
	httpd.log	Httpd resource log
	okerberos.log	Okerberos resource log
	oldap.log	OLdap resource log
	tomcat.log	Tomcat resource log
	send_alarm.log	Run log of the HA alarm sending script of the management node
timestamp	restart_stamp	NodeAgent start time log
tomcat	cas.log, localhost_access_cas_log.log	CAS run log
	catalina.log, catalina.out, host-manager.log, localhost.log, manager.log	Tomcat run log
	localhost_access_web_log.log	Log that records the access to REST APIs of FusionInsight Manager
	web.log	Run log of the web process
	northbound_ftp_sftp.log, snmp.log	Northbound log
watchdog	watchdog.log, feed_watchdog.log	watchdog run log
patch	oms_installPatch.log	OMS patch installation log
	agent_installPatch.log	Agent patch installation log
	agent_uninstallPatch.log	Agent patch uninstallation log
	NODE_AGENT_restoreFile.log	Agent patch restoration log

Type	Log File Name	Description
	NODE_AGENT_updateFile.log	Agent patch update log
	OMA_restoreFile.log	OMA patch restoration file log
	OMA_updateFile.log	OMA patch update file log
	CONTROLLER_restoreFile.log	CONTROLLER patch restoration file log
	CONTROLLER_updateFile.log	CONTROLLER patch update file log
	OMS_restoreFile.log	OMS patch restoration file log
	oms_uninstallPatch.log	OMS patch uninstallation log
	OMS_updateFile.log	OMS patch update file log
	createStackConf.log, decompress.log, decompress_OMS.log, distrExtractPatchOnOMS.log, slimReduction.log, switch_adapter.log	Patch installation log
sudo	sudo.log	Sudo script execution log

Log Levels

Table 10-24 describes the log levels provided by Manager. The priorities of log levels are FATAL, ERROR, WARN, INFO, and DEBUG in descending order. Logs whose levels are higher than or equal to the specified level are printed. The number of printed logs decreases as the specified log level increases.

Table 10-24 Log levels

Level	Description
FATAL	Logs of this level record fatal error information about the current event processing that may result in a system crash.
ERROR	Logs of this level record error information about the current event processing, which indicates that system running is abnormal.

Level	Description
WARN	Abnormal information about the current event processing. These abnormalities will not result in system faults.
INFO	Normal running status information about the system and events.
DEBUG	Logs of this level record the system information and system debugging information.

Log Formats

The following table lists the Manager log formats.

Table 10-25 Log formats

Type	Component	Format	Example
Controller, Httpd, logman, NodeAgent, okerberos, oldapserver, omm, tomcat, upgrade	Controller, Httpd, logman, NodeAgent, okerberos, oldapserver, omm, tomcat, upgrade	<*yyyy-MM-dd HH:mm:ss,SSS*>|<*Log level*>|<*Name of the thread that generates the log*>|<*Message in the log*>|<*Location where the log event occurs*>	2015-06-30 00:37:09,067 INFO [pool-1-thread-1] Completed Discovering Node. com.xxx.hadoop.om.controller.tasks.nodesetup.DiscoverNodeTask.execute(DiscoverNodeTask.java:299)

10.6.3 Viewing and Exporting Audit Logs

Scenario

This section describes how to view and export audit logs on MRS Manager. The audit logs can be used to trace security events, locate fault causes, and determine responsibilities.

The system record the following log information:

- User activity information, such as user login and logout, system user information modification, and system user group information modification
- User operation instruction information, such as cluster startup, stop, and software upgrade.

Procedure

- Viewing audit logs

a. On MRS Manager, click **Audit** to view the default audit logs.

   If the audit content of an audit log contains more than 256 characters, click the expand button of the audit log to expand the audit details. Click **Log File** to download the complete file and view the information.

   - By default, records are sorted in descending order by the **Occurred** column. You can click **Operation Type**, **Severity**, **Occurred**, **User**, **Host**, **Service**, **Instance**, or **Operation Result** to change the sorting mode.
   - All alarms of the same severity can be filtered by **Severity**. The results include cleared and uncleared alarms.

   Exported audit logs contain the following information:

   - **Sno**: indicates the number of audit logs generated by MRS Manager. The number is incremented by 1 when a new audit log is generated.
   - **Operation Type**: indicates the operation type of a user operation. There are nine scenarios: **Alarm**, **Auditlog**, **Backup And Restoration**, **Cluster**, **Collect Log**, **Host**, **Service**, **Tenant** and **User_Manager**. **User_Manager** is supported only in clusters with Kerberos authentication enabled. Each scenario contains different operation types. For example, **Alarm** includes **Export alarms**; **Cluster** includes **Start cluster**, and **Tenant** include **Add tenant**.
   - **Severity**: indicates the security level of each audit log, including **Critical**, **Major**, **Minor** and **Informational**.
   - **Start Time**: indicates the time when the operation starts. The time is .
   - **End Time**: indicates the time when the operation ends. The time is .
   - **User IP Address**: indicates the IP address used by a user to perform operations.
   - **User**: indicates the name of the user who performs the operation.
   - **Host**: indicates the node where the user operation is performed. The information is not saved if the operation does not involve a node.
   - **Service**: indicates the service in the cluster where the user operation is performed. The information is not saved if the operation does not involve a service.
   - **Instance**: indicates the role instance in the cluster where the user operation is performed. The information is not saved if the operation does not involve a role instance.
   - **Operation Result**: indicates the operation result, including **Successful**, **Failed** and **Unknown**.
   - **Content**: indicates execution information of the user operation.

b. Click **Advanced Search**. In the search area, set search criteria and click **Search** to view audit logs of the specified type. Click **Reset** to clear the search criteria.

NOTE

Start Time and **End Time** specify the start time and end time of the time range. You can search for alarms generated within the time range.

- Exporting audit logs
  a. In the audit log list, click **Export All** to export all logs.
  b. In the audit log list, select the check box of a log and click **Export** to export the log.

10.6.4 Exporting Service Logs

Scenario

This section describes how to export logs generated by each service role from MRS Manager.

Prerequisites

- You have obtained the access key ID (AK) and secret access key (SK) of the account.
- A parallel file system has been created in OBS.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 Click **Export Log** under **Maintenance**.

Step 3 Set a service for **Service**. Set **Host** to the IP address of the host where the service is deployed. Select the corresponding time for **Start Time** and **End Time**.

Step 4 In **Export To**, select a path for saving logs. This parameter is available only for clusters with Kerberos authentication enabled.

- **Local PC**: indicates that logs are saved to the local environment. Then go to **Step 8**.
- **OBS**: indicates that logs are saved to OBS. This is the default option. Then go to **Step 5**.

Step 5 Set **OBS Path** to the path for storing service logs on OBS.

The value must be a complete path and cannot start with a slash (/). The path can be nonexistent and will be automatically created by the system. The full path of OBS can contain a maximum of 900 bytes.

Step 6 In **Bucket**, enter the name of the created OBS file system.

Step 7 Set **AK** and **SK** to the access key ID and secret access key of the user.

Step 8 Click **OK**.

----End

10.6.5 Configuring Audit Log Dumping Parameters

Scenario

If audit logs are stored in the database for a long time, the disk space for the data directory may be insufficient. Therefore, you can set dump parameters to automatically dump audit logs to a specified directory on a server.

If you do not configure the audit log dumping, the system automatically saves the audit logs to a file when the number of audit logs reaches 100,000 pieces. The save path is **${BIGDATA_HOME}/OMSV100R001C00x8664/workspace/conf/data/operatelog** on the active management node. The file name format is **OperateLog_store_***YY_MM_DD_HH_MM_SS***.csv**. A maximum of 50 historical audit log files can be saved. The directory is automatically generated when audit logs are dumped for the first time.

Prerequisites

The ECS corresponding to the dump server must be in the same VPC as the master node of the MRS cluster, and the master node can access the IP address and specified port of the dump server. The SFTP service on the dump server is running properly.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 Choose **Dump Audit Log** under **Maintenance**.

Table 10-26 Audit log dump parameters

Parameter	Value	Description
Dump Audit Log	● On ● Off	(Mandatory) Specifies whether to enable audit log dumping. ● **On**: enables audit log dumping. ● **Off**: disables audit log dumping.
Dumping Mode	● By quantity ● By time	(Mandatory) Specifies the dump mode. ● **By quantity**: If the number of logs reaches the value of this parameter (**100,000** by default), the logs are dumped. ● **By time**: Logs are dumped at a specified date.

Parameter	Value	Description
SFTP IP	**192.168.10.51** (example value)	(Mandatory) Specifies the SFTP server for storing dumped audit logs.
SFTP Port	**22** (example value)	(Mandatory) Specifies the port of the SFTP server for storing dumped audit logs.
Save Path	**/opt/omm/oms/auditLog** (example value)	(Mandatory) Specifies the path for storing audit logs on the SFTP server.
SFTP Username	**root** (example value)	(Mandatory) Specifies the username for logging in to the SFTP server.
SFTP Password	**Root_123** (example value)	(Mandatory) Specifies the password for logging in to the SFTP server.
SFTP Public Key	-	(Optional) Specifies the public key of the SFTP server. You are advised to set the public key of the SFTP server. Otherwise, security risks may exist.
Dumping Date	November 06 (example value)	(Mandatory) Specifies the data when the system starts dumping audit logs. This parameter is valid when **Dump Mode** is set to **By time**. The logs to be dumped include all the audit logs generated before 00:00 on January 1 of the current year.

NOTE

Key fields in the audit log dump file are as follows:

- **USERTYPE** indicates the user type. Value **0** indicates the **Human-machine** user, and value **1** indicates the **Machine-machine** user.
- **LOGLEVEL** indicates the security level. Value **0** indicates critical, value **1** indicates major, value **2** indicates minor, and value **3** indicates informational.
- **OPERATERESULT** indicates the operation result. Value **0** indicates that the operation is successful, and value **1** indicates that the operation is failed.

----End

10.7 Health Check Management

10.7.1 Performing a Health Check

Scenario

To ensure that cluster parameters, configurations, and monitoring are correct and that the cluster can run stably for a long time, you can perform a health check during routine maintenance.

NOTE

A system health check includes MRS Manager, service-level, and host-level health checks:

- MRS Manager health checks focus on whether the unified management platform can provide management functions.
- Service-level health checks focus on whether components can provide services properly.
- Host-level health checks focus on whether host indicators are normal.

The system health check includes three types of check items: health status, related alarms, and customized monitoring indicators for each check object. The health check results are not always the same as the **Health Status** on the portal.

Procedure

- Manually perform the health check for all services.
  a. Click **Services** and select the target service.
  b. Choose **More** > **Start Service Health Check** to start the health check for the service.

  NOTE

  - The cluster health check includes Manager, service, and host status checks.
  - To perform cluster health checks, you can also choose **System** > **Check Health Check** > **Start Cluster Health Check** on MRS Manager.
  - To export the health check result, click **Export Report** in the upper left corner.
- Manually perform the health check for a service.
  a. Click **Services**. In the services list, click the desired service name.
  b. Choose **More** > **Start Service Health Check** to start the health check for the service.
- Manually perform the health check for a host.
  a. Click **Hosts**.
  b. Select the check box of the host for which you want to check the health status.
  c. Choose **More** > **Start Host Health Check** to start the health check for the host.
- Automatically performing a health check
  a. Click **System**.
  b. Click **Check Health Status** under **Maintenance**.
  c. Click **Configure Health Check** to configure automatic health check items.

     Periodic Health Check: specifies whether to enable automatic health check. The **Periodic Health Check** function is disabled by default. You can click to enable the function and select **Daily**, **Weekly**, or **Monthly** based on management requirements.

d. Click **OK** to save the settings. The **Health check configuration saved successfully** is displayed in the upper right corner.

10.7.2 Viewing and Exporting a Health Check Report

Scenario

You can view the health check result in MRS Manager and export the health check results for further analysis.

NOTE

A system health check includes MRS Manager, service-level, and host-level health checks:

- MRS Manager health checks focus on whether the unified management platform can provide management functions.
- Service-level health checks focus on whether components can provide services properly.
- Host-level health checks focus on whether host indicators are normal.

The system health check includes three types of check items: health status, related alarms, and customized monitoring indicators for each check object. The health check results are not always the same as the **Health Status** on the portal.

Prerequisites

You have performed a health check.

Procedure

Step 1 Click **Services**.

Step 2 Choose **More** > **View Cluster Health Check Report** to view the health check report of a cluster.

Step 3 Click **Export Report** on the health check report pane to export the report and view detailed information about check items.

NOTE

For details about how to rectify the faults of the check items, see **DBService Health Check Indicators** to **ZooKeeper Health Check Indicators**.

----End

10.7.3 Configuring the Number of Health Check Reports to Be Reserved

Scenario

Health check reports of MRS clusters, services, and hosts may vary with the time and scenario. You can modify the number of health check reports to be reserved on MRS Manager for later comparison.

This setting is valid for health check reports of clusters, services, and hosts. Report files are saved in **$BIGDATA_DATA_HOME/Manager/healthcheck** on the active management node by default and are automatically synchronized to the standby management node.

Prerequisites

Users have specified service requirements and planned the save time and health check frequency, and the disk space of the active and standby management nodes is sufficient.

Procedure

Step 1 Choose **System** > **Check Health Status** > **Configure Health Check**.

Step 2 Set **Max. Number of Health Check Reports** to the number of health check reports to be reserved. The value ranges from 1 to 100. The default value is 50.

Step 3 Click **OK** to save the settings. The **Health check configuration saved successfully** is displayed in the upper right corner.

----End

10.7.4 Managing Health Check Reports

Scenario

On MRS Manager, users can manage historical health check reports, for example, viewing, downloading, and deleting historical health check reports.

Procedure

- Download a specified health check report.
  a. Choose **System** > **Check Health Status**.
  b. Locate the row that contains the target health check report and click **Download** to download the report file.
- Download specified health check reports in batches.
  a. Choose **System** > **Check Health Status**.
  b. Select multiple health check reports and click **Download File** to download them.
- Delete a specified health check report.
  a. Choose **System** > **Check Health Status**.
  b. Locate the row that contains the target health check report and click **Delete** to delete the report file.
- Delete specified health check reports in batches.
  a. Choose **System** > **Check Health Status**.
  b. Select multiple health check reports and click **Delete File** to delete them.

10.7.5 DBService Health Check Indicators

Service Health Check

Indicator: Service Status

Description: This indicator is used to check whether the DBService service status is normal. If the status is abnormal, the service is unhealthy.

Handling method: If the indicator is abnormal, rectify the fault by referring to ALM-27001.

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check whether alarms exist on the host. If alarms exist, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

10.7.6 Flume Health Check Indicators

Service Health Status

Indicator: Service Status

Description: This indicator is used to check whether the Flume service status is normal. If the status is abnormal, the service is unhealthy.

Recovery Guide: If the indicator is abnormal, rectify the fault by referring to ALM-24000.

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check whether alarms exist on the host. If alarms exist, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

10.7.7 HBase Health Check Indicators

Normal RegionServer Count

Indicator: Normal RegionServer Count

Description: This indicator is used to check the number of RegionServers that are running properly in an HBase cluster.

Recovery Guide: If the indicator is abnormal, check whether the status of RegionServer is normal. If the status is abnormal, resolve the problem and check that the network is normal.

Service Health Status

Indicator: Service Status

Description: This indicator is used to check whether the HBase service status is normal. If the status is abnormal, the service is unhealthy.

Recovery Guide: If the indicator is abnormal, check whether the status of HMaster and RegionServer is normal. If the status is abnormal, resolve the problem. Then, check whether the status of the ZooKeeper service is faulty. On the HBase client, check whether the data in the HBase table can be correctly read and locate the data reading failure cause. Handle the alarm following instructions in the alarm processing document.

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check whether alarms exist. If alarms exist, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

10.7.8 Host Health Check Indicators

Swap Usage

Indicator: Swap Usage

Description: Swap usage of the system. The value is calculated using the following formula: Swap usage = Used swap size/Total swap size. Assume that the current threshold is set to 75.0%. If the usage of the file handles in the system exceeds the threshold, the system is unhealthy.

Recovery Guide:

1. Check the swap usage of the node.

   Log in to the unhealthy node and run the **free -m** command to check the total swap space and used swap space. If the swap space usage exceeds the threshold, go to **2**.
2. If the swap usage exceeds the threshold, you are advised to expand the system capacity, for example, add nodes.

Host File Handle Usage

Indicator: Host File Handle Usage

Description: This indicator indicates the file handle usage in the system. Host file handle usage = Number of used handles/Total number of handles. If the usage exceeds the threshold, the system is unhealthy.

Recovery Guide:

1. Check the file handle usage of the host.

Log in to the unhealthy node and run the **cat /proc/sys/fs/file-nr** command. In the command output, the first and third columns indicate the number of used handles and the total number of handles, respectively. If the usage exceeds the threshold, go to **2**.

2. If the file handle usage of the host exceeds the threshold, you are advised to check the system and analyze the file handle usage.

NTP Offset

Indicator: NTP Offset

Description: This indicator indicates the NTP time offset. If the time deviation exceeds the threshold, the system is unhealthy.

Recovery Guide:

1. Check the NTP time offset.

   Log in to the unhealthy node and run the **/usr/sbin/ntpq -np** command to view the information. In the command output, the **Offset** column indicates the time offset. If the time offset is greater than the threshold, go to **2**.

2. If the indicator is abnormal, check whether the clock source configuration is correct. Contact O&M personnel.

Average Load

Indicator: Average Load

Description: Average system load, indicating the average number of processes in the running queue in a specified period. The system average load is calculated using the load value obtained by the uptime command. Calculation method: (Load of 1 minute + Load of 5 minutes + Load of 15 minutes)/(3 x Number of CPUs). Assume that the current threshold is set to 2. If the average load exceeds 2, the system is unhealthy.

Recovery Guide:

1. Log in to the unhealthy node and run the **uptime** command. The last three columns in the command output indicate the load in 1 minute, 5 minutes, and 15 minutes, respectively. If the average system load exceeds the threshold, go to **2**.
2. If the system average load exceeds the threshold, you are advised to perform system capacity expansion, such as adding nodes.

D State Process

Indicator: D State Process

Description: This indicator indicates the unstoppable sleep process, that is, the process in the D state. A process that is in the D state is waiting for I/O, such as disk I/O and network I/O, and experiences an I/O exception. If any process in the D state exists in the system, the system is unhealthy.

Recovery Guide: If the indicator is abnormal, the system generates an alarm. You are advised to handle the alarm by referring to ALM-12028.

Hardware Status

Indicator: Hardware Status

Description: This indicator is used to check the system hardware status, including the CPU, memory, disk, power supply, and fan. This indicator obtains related hardware information using **ipmitool sdr elist**. If the hardware status is abnormal, the hardware is unhealthy.

Recovery Guide:

1. Log in to the node where the check result is unhealthy. Run the **ipmitool sdr elist** command to check system hardware status. The last column in the command output indicates the hardware status. If the status is included in the following fault description table, the check result is unhealthy.

Module	Symptom
Processor	IERR Thermal Trip FRB1/BIST failure FRB2/Hang in POST failure FRB3/Processor startup/init failure Configuration Error SM BIOS Uncorrectable CPU-complex Error Disabled Throttled Uncorrectable machine check exception
Power Supply	Failure detected Predictive failure Power Supply AC lost AC lost or out-of-range AC out-of-range, but present Config Error: Vendor Mismatch Config Error: Revision Mismatch Config Error: Processor Missing Config Error: Power Supply Rating Mismatch Config Error: Voltage Rating Mismatch Config Error

Module	Symptom
Power Unit	240VA power down Interlock power down AC lost Soft-power control failure Failure detected Predictive failure
Memory	Uncorrectable ECC Parity Memory Scrub Failed Memory Device Disabled Correctable ECC logging limit reached Configuration Error Throttled Critical Overtemperature
Drive Slot	Drive Fault Predictive Failure Parity Check In Progress In Critical Array In Failed Array Rebuild In Progress Rebuild Aborted
Battery	Low Failed

2. If the indicator is abnormal, contact O&M personnel.

Host Name

Indicator: Host Name

Description: This indicator is used to check whether the host name is set. If the host name is not set, the system is unhealthy. If the indicator is abnormal, you are advised to set the host name properly.

Recovery Guide:

1. Log in to the node where the check result is unhealthy.
2. Run the hostname host name command to change the host name to ensure that the host name is consistent with the planned host name.

   hostname*host name* For example, to change the host name to **Bigdata-OM-01**, run the **hostname Bigdata-OM-01** command.
3. Modify the host name configuration file.

Run the **vi /etc/HOSTNAME** command to edit the file. Change the file content to **Bigdata-OM-01**. Save the file, and exit.

Umask

Indicator: Umask

Description: This indicator is used to check whether the umask setting of user **omm** is correct. If Umask is not 0077, the system is unhealthy.

Recovery Guide:

1. If the indicator is abnormal, you are advised to set umask of user **omm** to 0077. Log in to the unhealthy node and run the **su - omm** command to switch to user **omm**.
2. Run the **vi ${BIGDATA_HOME}/.om_profile** command and change the value of **umask** to **0077**. Save and exit.

OMS HA Status

Indicator: OMS HA Status

Description: This indicator is used to check whether the OMS two-node cluster resources are normal. You can run the **${CONTROLLER_HOME}/sbin/status-oms.sh** command to view the detailed information about the status of the OMS two-node cluster resources. If any module is abnormal, the OMS is unhealthy.

Recovery Guide:

1. Log in to the active management node and run the **su - omm** command to switch to user **omm**. Run the **${CONTROLLER_HOME}/sbin/status-oms.sh** command to check the OMS status.
2. If floatip, okerberos, and oldap are abnormal, handle the problems by referring to ALM-12002, ALM-12004, and ALM-12005 respectively.
3. If other resources are abnormal, you are advised to view the logs of the faulty modules.

   If controller resources are abnormal, view **/var/log/Bigdata/controller/controller.log** of the faulty node.

   If CEP resources are abnormal, view **/var/log/Bigdata/omm/oms/cep/cep.log** of the faulty node.

   If AOS resources are abnormal, view **/var/log/Bigdata/controller/aos/aos.log** of the faulty node.

   If feed_watchdog resources are abnormal, view **/var/log/Bigdata/watchdog/watchdog.log** of the abnormal node.

   If HTTPD resources are abnormal, view **/var/log/Bigdata/httpd/error_log** of the abnormal node.

   If FMS resources are abnormal, view **/var/log/Bigdata/omm/oms/fms/fms.log** of the abnormal node.

   If PMS resources are abnormal, view **/var/log/Bigdata/omm/oms/pms/pms.log** of the abnormal node.

   If IAM resources are abnormal, view **/var/log/Bigdata/omm/oms/iam/iam.log** of the abnormal node.

If the GaussDB resource is abnormal, check the **/var/log/Bigdata/omm/oms/db/omm_gaussdba.log** of the abnormal node.

If NTP resources are abnormal, view **/var/log/Bigdata/omm/oms/ha/scriptlog/ha_ntp.log** of the abnormal node.

If Tomcat resources are abnormal, view **/var/log/Bigdata/tomcat/catalina.log** of the abnormal node.

4. If the fault cannot be rectified based on the logs, contact O&M personnel and send the collected fault logs.

Checking the Installation Directory and Data Directory

Indicator: Installation Directory and Data Directory Check

Description: This indicator checks the **lost+found** directory in the root directory of the disk partition where the installation directory (**/opt/Bigdata** by default) is located. If the directory contains the files of user **omm**, there are exceptions. When a node is abnormal, related files are stored in the **lost+found** directory. This indicator is used to check whether files are lost in such scenarios. Check the installation directory (for example, **/opt/Bigdata**) and data directory (for example, **/srv/BigData**). If any files of non-omm users exist in the two directories, the system is unhealthy.

Recovery Guide:

1. Log in to the unhealthy node and run the **su - omm** command to switch to user **omm**. Check whether files or folders of user omm exist in the **lost+found** directory.

   If the **omm** user file exists, you are advised to restore it and check again. If the **omm** user file does not exist, go to 2.

2. Check the installation directory and data directory. Check whether the files or folders of other users exist in the installation directory and data directory. If the files and folders are manually generated temporary files, you are advised to delete them and check again.

CPU Usage

Indicator: CPU Usage

Description: This indicator is used to check whether the CPU usage exceeds the threshold. If the disk usage exceeds the threshold, the system is unhealthy.

Recovery Guide: If the indicator is abnormal, the system generates an alarm. You are advised to handle the alarm by referring to ALM-12016.

Memory Usage

Indicator: Memory Usage

Description: This indicator is used to check whether the memory usage exceeds the threshold. If the disk usage exceeds the threshold, the system is unhealthy.

Recovery Guide: If the indicator is abnormal, the system generates an alarm. You are advised to handle the alarm by referring to ALM-12018.

Host Disk Usage

Indicator: Host Disk Usage

Description: This indicator is used to check whether the host disk usage exceeds the threshold. If the disk usage exceeds the threshold, the system is unhealthy.

Recovery Guide: If the indicator is abnormal, the system generates an alarm. You are advised to handle the alarm by referring to ALM-12017.

Host Disk Write Rate

Indicator: Host Disk Write Rate

Description: This indicator is used to check the disk write rate of a host. The write rate of the host disk may vary according to the service scenario. Therefore, the value of this indicator reflects only the specified value. You need to determine whether the indicator is normal in specified service scenarios.

Recovery Guide: Determine whether the current disk write rate is normal based on the service scenario.

Host Disk Read Rate

Indicator: Host Disk Read Rate

Description: This indicator is used to check the disk read rate of a host. The read rate of the host disk may vary by service scenario. Therefore, the value of this indicator reflects only the specified value. You need to determine whether the indicator is normal in specified service scenarios.

Recovery Guide: Determine whether the current disk read rate is normal based on the service scenario.

Host Service Plane Network Status

Indicator: Host Service Plane Network Status

Description: This indicator is used to check the connectivity of the service plane network of the cluster host. If the hosts are disconnected, the cluster is unhealthy.

Recovery Guide: If the single-plane networking is used, check the IP address of the single plane. For a dual-plane network, the operation procedure is as follows:

1. Check the network connectivity between the service plane IP addresses of the active and standby management nodes.

   If the network is abnormal, go to **3**.

   If the network is normal, go to **2**.

2. Check the network connectivity between the IP address of the active management node and the IP address of the abnormal node in the cluster.
3. If the network is disconnected, contact O&M personnel to rectify the network fault to ensure that the network meets service requirements.

Host Status

Indicator: Host Status

Description: This indicator is used to check whether the host status is normal. If a node is faulty, the host is unhealthy.

Recovery Guide: If the indicator is abnormal, rectify the fault by referring to ALM-12006.

Alarm Check

Indicator: Alarm Check

Description: This indicator is used to check whether alarms exist on the host. If alarms exist, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

10.7.9 HDFS Health Check Indicators

Average Packet Sending Time

Indicator: Average Packet Sending Time

Description: This indicator is used to collect statistics on the average time for the DataNode in the HDFS to execute SendPacket each time. If the average time is greater than 2,000,000 ns, the DataNode is unhealthy.

Recovery Guide: If the indicator is abnormal, check whether the network speed of the cluster is normal and whether the memory or CPU usage is too high. Check whether the HDFS load in the cluster is high.

Service Health Status

Indicator: Service Status

Description: This indicator is used to check whether the HDFS service status is normal. If a node is faulty, the host is unhealthy.

Recovery Guide: If the indicator is abnormal, check whether the health status of the KrbServer, LdapServer and ZooKeeper services are faulty. If yes, rectify the fault. Then, check whether the file writing failure is caused by HDFS SafeMode ON. Use the client to check whether data cannot be written into HDFS and locate the cause of the HDFS data writing failure. Handle the alarm following instructions in the alarm processing document.

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check whether alarms exist. If alarms exist, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

10.7.10 Hive Health Check Indicators

Maximum Number of Sessions Allowed by HiveServer

Indicator: Maximum Number of Sessions Allowed by HiveServer

Description: This indicator is used to check the maximum number of sessions that can be connected to Hive.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

Number of Sessions Connected to HiveServer

Indicator: Number of Sessions Connected to HiveServer

Description: This indicator is used to check the number of Hive connections.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

Service Health Status

Indicator: Service Status

Description: This indicator is used to check whether the Hive service status is normal. If the status is abnormal, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check whether alarms exist on the host. If alarms exist, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

10.7.11 Kafka Health Check Indicators

Number of Available Broker Nodes

Indicator: Number of Brokers

Description: This indicator is used to check the number of available Broker nodes in a cluster. If the number of available Broker nodes in a cluster is less than 2, the cluster is unhealthy.

Recovery Guide: If the indicator is abnormal, go to the Kafka service instance page and click the host name of the unavailable Broker instance. View the host health status in the **Overview** area. If the host health status is **Good**, rectify the fault by referring to the alarm handling suggestions in **Process Fault**. If the status is not **Good**, rectify the fault by referring to the handling procedure of the **Node Fault** alarm.

Service Health Status

Indicator: Service Status

Description: This indicator is used to check whether the Kafka service status is normal. If the status is abnormal, the service is unhealthy.

Recovery Guide: If the indicator is abnormal, rectify the fault by referring to the alarm "Kafka Service Unavailable".

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check whether alarms exist. If alarms exist, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

10.7.12 KrbServer Health Check Indicators

KerberosAdmin Service Availability

Indicator: KerberosAdmin Service Availability

Description: The system checks the KerberosAdmin service status. If the check result is abnormal, the KerberosAdmin service is unavailable.

Recovery Guide: If the indicator check result is abnormal, the possible cause is that the node where the KerberosAdmin service is located is faulty or the SlapdServer service is unavailable. During the KerberosAdmin service recovery, try the following operations:

1. Check whether the node where the KerberosAdmin service locates is faulty.
2. Check whether the SlapdServer service is unavailable.

KerberosServer Service Availability

Indicator: KerberosServer Service Availability

Description: The system checks the KerberosServer service status. If the check result is abnormal, the KerberosServer service is unavailable.

Recovery Guide: If the indicator check result is abnormal, the possible cause is that the node where the KerberosServer service is located is faulty or the SlapdServer service is unavailable. During the KerberosServer service recovery, try the following operations:

1. Check whether the node where the KerberosServer service locates is faulty.
2. Check whether the SlapdServer service is unavailable.

Service Health Status

Indicator: Service Status

Description: The system checks the KrbServer service status. If the check result is abnormal, the KrbServer service is unavailable.

Recovery Guide: If the indicator check result is abnormal, the possible cause is that the node where the KrbServer service resides is faulty or the LdapServer service is unavailable. For details, see the handling procedure of ALM-25500.

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check the alarm information about the KrbServer service. If any alarms exist, the KrbServer service may be abnormal.

Recovery Guide: If this indicator check result is abnormal, see the related alarm document to handle the alarms.

10.7.13 LdapServer Health Check Indicators

SlapdServer Service Availability

Indicator: SlapdServer Service Availability

Description: The system checks the SlapdServer service status. If the status is abnormal, the SlapdServer service is unavailable.

Recovery Guide: If the indicator check result is abnormal, the possible cause is that the node where the SlapdServer service is located is faulty or the SlapdServer process is faulty. During the SlapdServer service recovery, try the following operations:

1. Check whether the node where the SlapdServer service locates is faulty. For details, see ALM-12006.
2. Check whether the SlapdServer process is normal. For details, see ALM-12007.

Service Health Status

Indicator: Service Status

Description: This indicator is used to check the alarm information about the LdapServer service. If the status is abnormal, the LdapServer service is unavailable.

Recovery Guide: If the indicator check result is abnormal, the possible cause is that the node where the active LdapServer service resides is faulty or the active LdapServer process is faulty. For details, see ALM-25000.

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check the alarm information about the LdapServer service. If any alarms exist, the LdapServer service may be abnormal.

Recovery Guide: If this indicator check result is abnormal, see the related alarm document to handle the alarms.

10.7.14 Loader Health Check Indicators

ZooKeeper Health Status

Indicator: ZooKeeper health status

Description: This indicator is used to check whether the ZooKeeper health status is normal. If the status is abnormal, the ZooKeeper service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

HDFS Health Status

Indicator: HDFS health status

Description: This indicator is used to check whether the HDFS health status is normal. If the status is abnormal, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

DBService Health Status

Indicator: DBService Health Status

Description: This indicator is used to check whether the DBService health status is normal. If the status is abnormal, the DBService service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

Yarn Health Status

Indicator: Yarn health status

Description: This indicator is used to check whether the Yarn health status is normal. If the status is abnormal, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

MapReduce Health Status

Indicator: MapReduce Health Status

Description: This indicator is used to check whether the MapReduce health status is normal. If the status is abnormal, the MapReduce service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

Loader Process Status

Indicator: Loader Process Status

Description: This indicator is used to check whether the Loader process is normal. If the status is abnormal, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

Service Health Status

Indicator: Service Status

Description: This indicator is used to check whether the Loader service status is normal. If the status is abnormal, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check whether alarms exist for loader. If alarms exist, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

10.7.15 MapReduce Health Check Indicators

Service Health Status

Indicator: Service Status

Description: This indicator is used to check whether the MapReduce service status is normal. If the status is abnormal, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check whether alarms exist. If alarms exist, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

10.7.16 OMS Health Check Indicators

OMS Status Check

Indicator: OMS Status Check

Description: The OMS status check includes the HA status check and resource status check. The HA status includes **active**, **standby**, and **NULL**, indicating the active node, standby node, and unknown, respectively. The resource status includes normal, abnormal, and NULL. If the HA status is NULL, the HA status is unhealthy. If the resource status is NULL or abnormal, the resource status is unhealthy.

Table 10-27 OMS status description

Name	Description
HA state	**active**: indicates the active node. **standby**: indicates the standby node. **NULL**: unknown
Resource status	**normal**: All resources are normal. **abnormal**: indicates that abnormal resources exist. **NULL**: unknown

Recovery Guide:

1. Log in to the active management node and run the **su - omm** command to switch to user **omm**. Run the **${CONTROLLER_HOME}/sbin/status-oms.sh** command to check the status of OMS.
2. If the HA status is NULL, the system may be restarting. NULL is an intermediate state, and the HA status will automatically change to a normal state.
3. If the resource status is abnormal, certain component resources of FusionInsight Manager are abnormal. Check whether the status of components such as acs, aos cep, controller, feed_watchdog, fms, guassDB, httpd, iam, ntp, okerberos, oldap, pms, and tomcat component is normal.
4. If any Manager component resource is abnormal, see Manager component status check to rectify the fault.

Manager Component Status Check

Indicator: Manager Component Status Check

Description: This indicator is used to check the running status and HA status of Manager components. The resource running status includes **Normal** and **Abnormal**, and the resource HA status includes **Normal** and **Exception**. Manager components include Acs, Aos, Cep, Controller, feed_watchdog, Floatip, Fms, GaussDB, HeartBeatCheck, httpd, IAM, NTP, Okerberos, OLDAP, PMS, and Tomcat. If the running status and HA status is not Normal, the check result is unhealthy.

Table 10-28 Manager status description

Name	Description
Resource running status:	**Normal**: The system is running properly. **Abnormal**: The running is abnormal. **Stopped**: The task is stopped. **Unknown**: The status is unknown. **Starting**: The process is being started. **Stopping**: The task is being stopped. **Active_normal**: The active node is running properly. **Standby_normal**: The standby node is running properly. **Raising_active**: The node is being promoted to be the active node. **Lowing_standby**: The node is being set to be the standby node. **No_action**: the action does not exist. **Repairing**: The disk is being repaired. **NULL**: unknown
Resource HA status	**Normal**: the status is normal. **Exception**: indicates a fault. **Non_steady**: indicates the non-steady state. **Unknown**: unknown **NULL**: unknown

Recovery Guide:

1. Log in to the active management node and run the **su - omm** command to switch to user **omm**. Run the **${CONTROLLER_HOME}/sbin/status-oms.sh** command to check the status of OMS.
2. If floatip, okerberos, and oldap are abnormal, handle the problems by referring to ALM-12002, ALM-12004, and ALM-12005 respectively.
3. If other resources are abnormal, you are advised to view the logs of the faulty modules.

   If controller resources are abnormal, view **/var/log/Bigdata/controller/controller.log** of the faulty node.

   If CEP resources are abnormal, view **/var/log/Bigdata/omm/oms/cep/cep.log** of the faulty node.

   If AOS resources are abnormal, view **/var/log/Bigdata/controller/aos/aos.log** of the faulty node.

   If feed_watchdog resources are abnormal, view **/var/log/Bigdata/watchdog/watchdog.log** of the abnormal node.

If HTTPD resources are abnormal, view **/var/log/Bigdata/httpd/error_log** of the abnormal node.

If FMS resources are abnormal, view **/var/log/Bigdata/omm/oms/fms/fms.log** of the abnormal node.

If PMS resources are abnormal, view **/var/log/Bigdata/omm/oms/pms/pms.log** of the abnormal node.

If IAM resources are abnormal, view **/var/log/Bigdata/omm/oms/iam/iam.log** of the abnormal node.

If the GaussDB resource is abnormal, check the **/var/log/Bigdata/omm/oms/db/omm_gaussdba.log** of the abnormal node.

If NTP resources are abnormal, view **/var/log/Bigdata/omm/oms/ha/scriptlog/ha_ntp.log** of the abnormal node.

If Tomcat resources are abnormal, view **/var/log/Bigdata/tomcat/catalina.log** of the abnormal node.

4. If the fault cannot be rectified based on the logs, contact O&M personnel and send the collected fault logs.

OMA Running Status

Indicator: OMA Running Status

Description: This indicator is used to check the running status of the OMA. The status can be **Running** or **Stopped**. If the OMA is **Stopped**, the OMA is unhealthy.

Recovery Guide:

1. Log in to the unhealthy node and run the **su - omm** command to switch to user **omm**.
2. Run **${OMA_PATH}/restart_oma_app** to manually start the OMA and check again. If the check result is still unhealthy, go to **3**.
3. If manually starting the OMA cannot resolve the problem, you are advised to check the OMA logs in **/var/log/Bigdata/omm/oma/omm_agent.log**.
4. If the fault cannot be rectified based on the logs, contact O&M personnel and send the collected fault logs.

SSH Trust Between Each Node and the Active Management Node

Indicator: SSH Trust Between Each Node and the Active Management Node

Description: This indicator is used to check whether the SSH mutual trust is normal. If you can switch to another node through SSH from the active OMS node as user omm without the need of entering the password, SSH communication is normal. Otherwise, SSH communication is abnormal. In addition, if you can switch to another node through SSH from the active OMS node but fail to switch to the active OMS node from the other nodes, SSH communication is abnormal.

Recovery Guide:

1. If the indicator check result is abnormal, the SSH trust relationships between the nodes and the active management node are abnormal. In this case, check whether the permission of the **/home/omm** directory is **omm**. If non-omm users have the directory permission, the SSH trust relationship may be

abnormal. You are advised to run **chown omm:wheel** to modify the permission and check again. If the permission on the **/home/omm** directory is normal, go to **2**.

2. The SSH trust relationship exception may cause heartbeat exceptions between Controller and NodeAgent, resulting in node fault alarms. In this case, rectify the fault by referring to the handling procedure of ALM-12006.

Process Running Time

Indicator: Running Time of NodeAgent, Controller, and Tomcat

Description: This indicator is used to check the running time of the NodeAgent, Controller, and Tomcat processes. If the time is less than half an hour (1,800s), the process may have been restarted. You are advised to check the process after half an hour. If multiple check results indicate that the process runs for less than half an hour, the process is abnormal.

Recovery Guide:

1. Log in to the unhealthy node and run the **su - omm** command to switch to user **omm**.
2. Run the following command to check the PID based on the process name:

   ps -ef | grep NodeAgent
3. Run the following command to check the process startup time based on the PID:

   ps -p pid -o lstart
4. Check whether the process start time is normal. If the process restarts repeatedly, go to **5**.
5. View the related logs and analyze restart causes.

   If the runtime of NodeAgent is abnormal, check **/var/log/Bigdata/nodeagent/agentlog/agent.log**.

   If the Controller running time is abnormal, check the **/var/log/Bigdata/controller/controller.log** file.

   If the Tomcat running time is abnormal, check the **/var/log/Bigdata/tomcat/web.log** file.
6. If the fault cannot be rectified based on the logs, contact O&M personnel and send the collected fault logs.

Account and Password Expiration Check

Indicator: Account and Password Expiration Check

Description: This indicator checks the two operating system users **omm** and **ommdba** of MRS. For OS users, both the account and password expiration time must be checked. If the validity period of the account or password is not greater than 15 days, the account is abnormal.

Recovery Guide: If the validity period of the account or password is less than or equal to 15 days, contact O&M personnel.

10.7.17 Spark Health Check Indicators

Service Health Status

Indicator: Service Status

Description: This indicator is used to check whether the Spark service status is normal. If the status is abnormal, the service is unhealthy.

Recovery Guide: If the indicator is abnormal, rectify the fault by referring to ALM-28001.

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check whether alarms exist. If alarms exist, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

10.7.18 Storm Health Check Indicators

Number of Working Nodes

Indicator: Number of Supervisors

Description: This indicator is used to check the number of available Supervisors in a cluster. If the number of available Supervisors in a cluster is less than 1, the cluster is unhealthy.

Recovery Guide: If the indicator is abnormal, go to the Streaming service instance page and click the host name of the unavailable Supervisor instance. View the host health status in the **Overview** area. If the host health status is **Good**, rectify the fault by referring to ALM-12007 Process Faults. If the status is not **Good**, rectify the fault by referring to the handling procedure of the ALM-12006 Node Faults.

Number of Idle Slots

Indicator: Number of Idle Slots

Description: This indicator is used to check the number of idle slots in a cluster. If the number of idle slots in a cluster is less than 1, the cluster is unhealthy.

Recovery Guide: If the indicator is abnormal, go to the Storm service instance page and check the health status of the Supervisor instance. If the health status of all Supervisor instances is **Good**, you need to expand the capacity of the Core node in the cluster. If not, rectify the fault by referring to ALM-12007 Process Faults.

Service Health Status

Indicator: Service Status

Description: This indicator is used to check whether the Storm service status is normal. If the status is abnormal, the service is unhealthy.

Recovery Guide: If the indicator is abnormal, rectify the fault by referring to the alarm "ALM-26051 Storm Service Unavailable".

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check whether alarms exist. If alarms exist, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

10.7.19 Yarn Health Check Indicators

Service Health Status

Indicator: Service Status

Description: This indicator is used to check whether the Yarn service status is normal. If the number of NodeManager nodes cannot be obtained, the system is unhealthy.

Recovery Guide: If this indicator is abnormal, you can handle the alarm by referring to the alarm handling guide and make sure that the network is normal.

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check whether alarms exist. If alarms exist, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

10.7.20 ZooKeeper Health Check Indicators

Average ZooKeeper Request Processing Latency

Indicator: Average ZooKeeper Service Request Processing Latency

Description: This indicator is used to check the average delay for the ZooKeeper service to process requests. If the average delay is greater than 300 ms, the ZooKeeper service is unhealthy.

Recovery Guide: If the indicator is abnormal, check whether the network speed of the cluster is normal and whether the memory or CPU usage is too high.

ZooKeeper Connections Usage

Indicator: ZooKeeper Connections Usage

Description: This indicator is used to check whether the ZooKeeper memory usage exceeds 80%. If the disk usage exceeds the threshold, the system is unhealthy.

Recovery Guide: If the indicator is abnormal, you are advised to increase the memory available for the ZooKeeper service. The method of increasing the memory is as follows: Increase the value of **-Xmx** in the **GC_OPTS** configuration item in the ZooKeeper service. After the modification, restart the ZooKeeper service for the configuration to take effect.

Service Health Status

Indicator: Service Status

Description: This indicator is used to check whether ZooKeeper service status is normal. If the status is abnormal, the service is unhealthy.

Recovery Guide: If the indicator is abnormal, check whether the health status of the KrbServer and LdapServer services is faulty. If yes, rectify the fault. Log in to the ZooKeeper client, check whether the ZooKeeper data writing fails. If yes, find the failure cause based on the error message and handle the fault according to error message. Rectify the fault by following the procedure for handling ALM-13000.

Alarm Check

Indicator: Alarm Information

Description: This indicator is used to check whether alarms exist. If alarms exist, the service is unhealthy.

Recovery Guide: If this indicator is abnormal, you can rectify the fault by referring to the alarm handling guide.

10.8 Static Service Pool Management

10.8.1 Viewing the Status of a Static Service Pool

Scenario

MRS Manager manages and isolates service resources that are not running on YARN through the static service resource pool. It dynamically manages the total CPU, I/O, and memory resources that can be used by HDFS and YARN on the deployment node. The system supports time-based automatic adjustment of static service resource pools. This enables the cluster to automatically adjust the parameter values at different periods to ensure more efficient resource utilization.

On MRS Manager, you can view the monitoring metrics of the resources used by each service in the static service pool. The monitoring metrics are as follows:

- Service Total CPU Usage
- Service Total Disk I/O Read Speed
- Service Total Disk I/O Write Speed

- Service Total Memory Usage

Procedure

Step 1 On MRS Manager, click **System**. In the **Resource** area, click **Configure Static Service Pool**.

Step 2 Click **Status**.

Step 3 Check the system resource adjustment base values.

- **System Resource Adjustment Base** indicates the maximum volume of resources that can be used by each node in the cluster. If a node has only one service, the service exclusively occupies the available resources on the node. If a node has multiple services, all services share the available resources on the node.
- **CPU(%)** indicates the maximum number of CPUs that can be used by services on a node.
- **Memory(%)** indicates the maximum memory that can be used by services on a node.

Step 4 Check the cluster service resource usage.

In the chart area, select **All services** from the service drop-down list box. The resource usage status of all services in the service pool is displayed.

NOTE

Effective Configuration Group indicates the resource control configuration group used by the cluster service. By default, the **default** configuration group is used at all time every day, indicating that the cluster service can use all CPUs and 70% memory of the node.

Step 5 View the resource usage of a single service.

In the chart area, select a service from the service drop-down list box. The resource usage status of the service is displayed.

Step 6 You can set the interval for automatically refreshing the page.

The following refresh interval options are supported:

- **Refresh every 30 seconds**
- **Refresh every 60 seconds**
- **Stop refreshing**

Step 7 In the **Period** area, select a time range for viewing service resources. The options are as follows:

- Real time
- Last 3 hours
- Last 6 hours
- Last 24 hours
- Last week
- Last month
- Last 3 months

- Last 6 months
- Customize: If you select this option, you can customize the period for viewing monitoring data.

Step 8 Click **View** to view the service resource data in the corresponding time range.

Step 9 Customize a service resource report.

1. Click **Customize** and select the service source indicators to be displayed.
   - Service Total Disk I/O Read Speed
   - Service Total Memory Usage
   - Service Total Disk I/O Write Speed
   - Service Total CPU Usage
2. Click **OK** to save the selected monitoring metrics for display.

   NOTE

   Click **Clear** to cancel all the selected monitoring metrics in a batch.

Step 10 Export a monitoring report.

Click **Export**. MRS Manager will generate a report about the selected service resources in a specified time of period. Save the report.

NOTE

To view the curve charts of monitoring metrics in a specified period, click **View**.

----End

10.8.2 Configuring a Static Service Pool

Scenario

If you need to control the node resources that can be used by the cluster service or the CPU usage of the node used by the cluster in different time periods, you can adjust the resource base on MRS Manager and customize the resource configuration groups.

Prerequisites

- After the static service pool is configured, the HDFS and YARN services need to be restarted. During the restart, the services are unavailable.
- After a static service pool is configured, the maximum number of resources used by each service and role instance cannot exceed the upper limit.

Procedure

Step 1 Modify the system resource adjustment base.

1. On MRS Manager, click **System**. In the **Resource** area, click **Configure Static Service Pool**.
2. Click **Configuration**. The service pool configuration group management page is displayed.

3. In the **System Resource Adjustment Base** area, change the values of **CPU(%)** and **Memory(%)** .

   Modifying **System Resource Adjustment Base** limits the maximum physical CPU and memory resource percentage of nodes that can be used by the Flume, HBase, HDFS, Impala and YARN services. If multiple services are deployed on the same node, the maximum physical resource usage of all services cannot exceed the adjusted CPU or memory usage.

4. Click **Next**.

   If you need to modify the parameters again, click **Previous** in the lower part of the page.

Step 2 Modify the **default** configuration group of the service pool.

1. Click **default**. In the **Service Pool Configuration** table, set **CPU LIMIT(%)**, **CPU SHARE(%)**, **I/O(%)**, and **Memory(%)** for the Flume, HBase, HDFS, Impala and YARN services.

NOTE

- The sum of **CPU LIMIT(%)** used by all services can exceed 100%.
- The sum of **CPU SHARE(%)** and **I/O(%)** used by all services must be 100%. For example, if CPU resources are allocated to the HDFS and Yarn services, the total CPU resources allocated to the two services are 100%.
- The sum of **Memory(%)** used by all services can be greater than, smaller than, or equal to 100%.
- **Memory(%)** cannot take effect dynamically and can only be modified in the default configuration group.

2. Click in the blank area of the page to complete the editing. MRS Manager generates the correct values of service pool parameters in the **Detailed Configuration** area based on the cluster hardware resources and allocation information.

3. You can click ✎ on the right of **Detailed Configuration** to modify the parameter values of the service pool based on service requirements.

   In the **Service Pool Configuration** area, click the specified service name. The **Detailed Configuration** area displays only the parameters of the service. Manual changing of parameter values does not refresh the service resource usage. In added configuration groups, the configuration group numbers of the parameters that take effect dynamically will be displayed. For example, **HBase: RegionServer: dynamic-config1.RES_CPUSET_PERCENTAGE**. The parameter functions do not change.

Table 10-29 Parameters of the static service pool

Parameter	Description
– RES_CPUSET_PERCENTAGE – dynamic-configX.RES_CPUSET_PERCENTAGE	Configures the service CPU percentage.

Parameter	Description
– RES_CPU_SHARE – dynamic-configX.RES_CPU_SHARE	Configures the service CPU share.
– RES_BLKIO_WEIGHT – dynamic-configX.RES_BLKIO_WEIGHT	Configures service I/O usage.
HBASE_HEAPSIZE	Configures the maximum JVM memory for RegionServer.
HADOOP_HEAPSIZE	Configures the maximum JVM memory of a DataNode.
yarn.nodemanager.resource.memory-mb	Configures the memory that can be used by NodeManager on the current node.
dfs.datanode.max.locked.memory	Configures the maximum memory that can be used by a DataNode as the HDFS cache.
FLUME_HEAPSIZE	Configures the maximum JVM memory that can be used by each Flume instance.
IMPALAD_MEM_LIMIT	Configures the maximum memory that can be used by an Impalad instance.

Step 3 Add a customized resource configuration group.

1. Determine whether to automatically adjust resource configurations based on the time.

   If yes, go to **Step 3.2**.

   If no, go to **Step 4**.

2. Click ⊕ to add a resource configuration group. In the **Scheduling Time** area, click ✎. The time policy configuration page is displayed.

   Modify the following parameters based on service requirements and click **OK**.

   - **Repeat**: If selected, the resource configuration group runs repeatedly based on the scheduling period. If not selected, set the date and time when the configuration of the group of resources can be applied.
   - **Repeat Policy**: can be set to **Daily**, **Weekly**, and **Monthly**. This parameter is valid only when **Repeat** is selected.
   - **Between**: indicates the time period between the start time and end time when the resource configuration is applied. Set a unique time range. If the time range overlaps with that of an existing group of resource configuration, the time range cannot be saved. This parameter is valid only when **Repeat** is selected.

> NOTE
>
> - The **default** group of resource configuration takes effect in all undefined time segments.
> - The newly added resource group is a parameter set that takes effect dynamically in a specified time range.
> - The newly added resource group can be deleted. A maximum of four resource configuration groups that take effect dynamically can be added.
> - Select a repetition policy. If the end time is earlier than the start time, the next day is labeled by default. For example, if a validity period ranges from 22:00 to 06:00, the customized resource configuration takes effect from 22:00 on the current day to 06:00 on the next day.
> - If the repeat policy types of multiple configuration groups are different, the time ranges can overlap. The policy types are listed as follows by priority from low to high: daily, weekly, and monthly. The following is an example. There are two resource configuration groups using the monthly and daily policies, respectively. Their application time ranges in a day overlap as follows: [04:00 to 07:00] and [06:00 to 08:00]. In this case, the configuration of the group that uses the monthly policy prevails.
> - If the repeat policy types of multiple resource configuration groups are the same, the time ranges of different dates can overlap. For example, if there are two weekly scheduling groups, you can set the same time range on different day for them, such as to 04:00 to 07:00, on Monday and Wednesday, respectively.

3. On the **Service Pool Configuration** page, modify the resource configuration of each service. Click the blank area on the page to complete the editing, and go to **Step 4**.

   You can click ✎ on the right of **Service Pool Configuration** to modify the parameters. Click ✎ in the **Detailed Configuration** area to manually update the parameter values generated by the system based on service requirements.

Step 4 Saves the settings.

Click **Save**. In the **Save Configuration** dialog box, select **Restart the affected services or instances**. Click **OK** to save the settings and restart related services.

Operation succeeded is displayed. click **Finish**. The service is started successfully.

----End

10.9 Tenant Management

10.9.1 Overview

Definition

An MRS cluster provides various resources and services for multiple organizations, departments, or applications to share. The cluster provides tenants as a logical entity to use these resources and services. A mode involving different tenants is called multi-tenant mode. Currently, only the analysis cluster supports tenant management.

Principles

The MRS cluster provides the multi-tenant function. It supports a layered tenant model and allows dynamic adding or deleting of tenants to isolate resources. It dynamically manages and configures tenants' computing and storage resources.

The computing resources indicate tenants' Yarn task queue resources. The task queue quota can be modified, and the task queue usage status and statistics can be viewed.

The storage resources can be stored on HDFS. You can add and delete the HDFS storage directories of tenants, and set the quotas of file quantity and the storage space of the directories.

As the unified tenant management platform of MRS clusters, MRS Manager provides enterprises with time-tested multi-tenant management models, enabling centralized tenant and service management. Tenants can create and manage tenants in a cluster based on service requirements.

- Roles, computing resources, and storage resources are automatically created when tenants are created. By default, all permissions of the new computing resources and storage resources are allocated to a tenant's roles.
- Permissions to view the current tenant's resources, add a subtenant, and manage the subtenant's resources are granted to the tenant's roles by default.
- After you have modified the tenant's computing or storage resources, permissions of the tenant's roles are automatically updated.

MRS Manager supports a maximum of 512 tenants. The tenants that are created by default in the system contain **default**. Tenants that are in the topmost layer with the default tenant are called level-1 tenants.

Resource Pools

Yarn task queues support only the label-based scheduling policy. This policy enables Yarn task queues to associate NodeManagers that have specific node labels. In this way, Yarn tasks run on specified nodes so that tasks are scheduled and certain hardware resources are utilized. For example, Yarn tasks requiring a large memory capacity can run on nodes with a large memory capacity by means of label association, preventing poor service performance.

In an MRS cluster, the tenant logically divides Yarn cluster nodes to combine multiple NodeManagers into a resource pool. Yarn task queues can be associated with specified resource pools by configuring queue capacity policies, ensuring efficient and independent resource utilization in the resource pools.

MRS Manager supports a maximum of 50 resource pools. The system has a **Default** resource pool.

10.9.2 Creating a Tenant

Scenario

You can create a tenant on MRS Manager to specify the resource usage.

Prerequisites

- A tenant name has been planned. The name must not be the same as that of a role or Yarn queue that exists in the current cluster.
- If a tenant requires storage resources, a storage directory has been planned based on service requirements, and the planned directory does not exist under the HDFS directory.
- The resources that can be allocated to the current tenant have been planned and the sum of the resource percentages of direct sub-tenants under the parent tenant at every level does not exceed 100%.

Procedure

Step 1 On MRS Manager, click **Tenant**.

Step 2 Click **Create Tenant**. On the page that is displayed, configure tenant properties.

Table 10-30 Tenant parameters

Parameter	Description
Name	Specifies the name of the current tenant. The value consists of 3 to 20 characters, and can contain letters, digits, and underscores (_).
Tenant Type	The options include **Leaf** and **Non-leaf**. If **Leaf** is selected, the current tenant is a leaf tenant and no sub-tenant can be added. If **Non-leaf** is selected, sub-tenants can be added to the current tenant.
Dynamic Resources	Specifies the dynamic computing resources for the current tenant. The system automatically creates a task queue named after the tenant name in Yarn. When dynamic resources are not **Yarn**, the system does not automatically create a task queue.
Default Resource Pool Capacity (%)	Specifies the percentage of the computing resources used by the current tenant in the **default** resource pool.
Default Resource Pool Max. Capacity (%)	Specifies the maximum percentage of the computing resources used by the current tenant in the **default** resource pool.
Storage Resource	Specifies storage resources for the current tenant. The system automatically creates a file folder named after the tenant name in the **/tenant** directory. When a tenant is created for the first time, the system automatically creates the **/tenant** directory in the HDFS root directory. If storage resources are not **HDFS**, the system does not create a storage directory under the root directory of HDFS.

Parameter	Description
Space Quota (MB)	Specifies the quota for HDFS storage space used by the current tenant. The value ranges from **1** to **8796093022208**. The unit is MB. This parameter indicates the maximum HDFS storage space that can be used by a tenant, but does not indicate the actual space used. If the value is greater than the size of the HDFS physical disk, the maximum space available is the full space of the oHDFS physical disk. **NOTE** To ensure data reliability, one copy of a file is automatically generated when the file is stored in HDFS. That is, two copies of the same file are stored by default. The HDFS storage space indicates the total disk space occupied by all these copies. For example, if the value of **Storage Space Quota** is set to **500**, the actual space for storing files is about 250 MB (500/2 = 250).
Storage Path	Specifies the tenant's HDFS storage directory. The system automatically creates a file folder named after the tenant name in the **/tenant** directory by default. For example, the default HDFS storage directory for tenant **ta1** is **tenant/ta1**. When a tenant is created for the first time, the system automatically creates the **/tenant** directory in the HDFS root directory. The storage path is customizable.
Service	Specifies other service resources associated with the current tenant. HBase is supported. To configure this parameter, click **Associate Services**. In the dialog box that is displayed, set **Service** to **HBase**. If **Association Mode** is set to **Exclusive**, service resources are occupied exclusively. If **share** is selected, service resources are shared.
Description	Specifies the description of the current tenant.

Step 3 Click **OK** to save the settings.

It takes a few minutes to save the settings. If the **Tenant created successfully** is displayed in the upper-right corner, the tenant is added successfully.

NOTE

- Roles, computing resources, and storage resources are automatically created when tenants are created.
- The new role has permissions on the computing and storage resources. The role and its permissions are controlled by the system automatically and cannot be controlled manually under **Manage Role**.
- If you want to use the tenant, create a system user and assign the Manager_tenant role and the role corresponding to the tenant to the user. For details, see **Creating a User**.

----End

Related Tasks

Viewing an added tenant

Step 1 On MRS Manager, click **Tenant**.

Step 2 In the tenant list on the left, click the name of the added tenant.

The **Summary** tab is displayed on the right by default.

Step 3 View **Basic Information**, **Resource Quota**, and **Statistics** of the tenant.

If HDFS is in the **Stopped** state, **Available** and **Used** of **Space** in **Resource Quota** are **unknown.**

----End

10.9.3 Creating a Sub-tenant

Scenario

You can create a sub-tenant on MRS Manager if the resources of the current tenant need to be further allocated.

Prerequisites

- A parent tenant has been added.
- A tenant name has been planned. The name must not be the same as that of a role or Yarn queue that exists in the current cluster.
- If a sub-tenant requires storage resources, a storage directory has been planned based on service requirements, and the planned directory does not exist under the storage directory of the parent tenant.
- The resources that can be allocated to the current tenant have been planned and the sum of the resource percentages of direct sub-tenants under the parent tenant at every level does not exceed 100%.

Procedure

Step 1 On MRS Manager, click **Tenant**.

Step 2 In the tenant list on the left, move the cursor to the tenant node to which a sub-tenant is to be added. Click **Create sub-tenant**. On the displayed page, configure the sub-tenant attributes according to the following table:

Table 10-31 Sub-tenant parameters

Parameter	Description
Parent tenant	Specifies the name of the parent tenant.
Name	Specifies the name of the current tenant. The value consists of 3 to 20 characters, and can contain letters, digits, and underscores (_).

Parameter	Description
Tenant Type	The options include **Leaf** and **Non-leaf**. If **Leaf** is selected, the current tenant is a leaf tenant and no sub-tenant can be added. If **Non-leaf** is selected, sub-tenants can be added to the current tenant.
Dynamic Resources	Specifies the dynamic computing resources for the current tenant. The system automatically creates a task queue named after the sub-tenant name in the Yarn parent queue. When dynamic resources are not **Yarn**, the system does not automatically create a task queue. If the parent tenant does not have dynamic resources, the sub-tenant cannot use dynamic resources.
Default Resource Pool Capacity (%)	Specifies the percentage of the resources used by the current tenant. The base value is the total resources of the parent tenant.
Default Resource Pool Max. Capacity (%)	Specifies the maximum percentage of the computing resources used by the current tenant. The base value is the total resources of the parent tenant.
Storage Resource	Specifies storage resources for the current tenant. The system automatically creates a file in the HDFS parent tenant directory. The file is named the same as the name of the sub-tenant. If storage resources are not **HDFS**, the system does not create a storage directory under the root directory of HDFS. If the parent tenant does not have storage resources, the sub-tenant cannot use storage resources.
Space Quota (MB)	Specifies the quota for HDFS storage space used by the current tenant. The minimum value is 1, and the maximum value is the total storage quota of the parent tenant. The unit is MB. This parameter indicates the maximum HDFS storage space that can be used by a tenant, but does not indicate the actual space used. If the value is greater than the size of the HDFS physical disk, the maximum space available is the full space of the oHDFS physical disk. If the quota is greater than the quota of the parent tenant, the actual storage capacity is subject to the quota of the parent tenant. **NOTE** To ensure data reliability, one copy of a file is automatically generated when the file is stored in HDFS. That is, two copies of the same file are stored by default. The HDFS storage space indicates the total disk space occupied by all these copies. For example, if the value is set to **500**, the actual space for storing files is about 250 MB (500/2 = 250).

Parameter	Description
Storage Path	Specifies the tenant's HDFS storage directory. The system automatically creates a file folder named after the sub-tenant name in the directory of the parent tenant by default. For example, if the sub-tenant is **ta1s** and the parent directory is **tenant/ta1**, the system sets this parameter for the sub-tenant to **tenant/ta1/ta1s**. The storage path is customizable in the parent directory. The parent directory for the storage path must be the storage directory of the parent tenant.
Service	Specifies other service resources associated with the current tenant. HBase is supported. To configure this parameter, click **Associate Services**. In the dialog box that is displayed, set **Service** to **HBase**. If **Association Mode** is set to **Exclusive**, service resources are occupied exclusively. If **share** is selected, service resources are shared.
Description	Specifies the description of the current tenant.

Step 3 Click **OK** to save the settings.

It takes a few minutes to save the settings. If the **Tenant created successfully** is displayed in the upper-right corner, the tenant is added successfully. The tenant is created successfully.

NOTE

- Roles, computing resources, and storage resources are automatically created when tenants are created.
- The new role has permissions on the computing and storage resources. The role and its permissions are controlled by the system automatically and cannot be controlled manually under **Manage Role**.
- When using this tenant, create a system user and assign the user a related tenant role. For details, see **Creating a User**.

----End

10.9.4 Deleting a tenant

Scenario

You can delete a tenant that is not required on MRS Manager.

Prerequisites

- A tenant has been added.
- You have checked whether the tenant to be deleted has sub-tenants. If the tenant has sub-tenants, delete them; otherwise, you cannot delete the tenant.

- The role of the tenant to be deleted cannot be associated with any user or user group. For details about how to cancel the binding between a role and a user, see **Modifying User Information**.

Procedure

Step 1 On MRS Manager, click **Tenant**.

Step 2 In the tenant list on the left, move the cursor to the tenant node to be deleted and click **Delete**.

The **Delete Tenant** dialog box is displayed. If you want to save the tenant data, select **Reserve the data of this tenant**. Otherwise, the tenant's storage space will be deleted.

Step 3 Click OK to save the settings.

It takes a few minutes to save the configuration. After the tenant is deleted successfully, the role and storage space of the tenant are also deleted.

NOTE

- After the tenant is deleted, the task queue of the tenant still exists in Yarn.
- If you choose not to reserve data when deleting the parent tenant, data of sub-tenants is also deleted if the sub-tenants use storage resources.

----End

10.9.5 Managing a Tenant Directory

Scenario

You can manage the HDFS storage directory used by a specific tenant on MRS Manager. The management operations include adding a tenant directory, modifying the directory file quota, modifying the storage space, and deleting a directory.

Prerequisites

A tenant associated with HDFS storage resources has been added.

Procedure

- Viewing a tenant directory
  a. On MRS Manager, click **Tenant**.
  b. In the tenant list on the left, click the target tenant.
  c. Click the **Resource** tab.
  d. View the **HDFS Storage** table.
     - The Quota column indicates the quantity quotas of files and directories.
     - The **Storage Space Quota** column indicates the storage space size of the tenant directory.

- Adding a tenant directory
  1. On MRS Manager, click **Tenant**.
  2. In the tenant list on the left, click the tenant whose HDFS storage directory needs to be added.
  3. Click the **Resource** tab.
  4. In the **HDFS Storage** table, click **Create Directory**.
     - In **Parent Directory**, select a storage directory of a parent tenant.

       This parameter applies only to sub-tenants. If the parent tenant has multiple directories, select any of them.
     - Set **Path** to a tenant directory path.

       NOTE

       - If the current tenant is not a sub-tenant, the new path is created in the HDFS root directory.
       - If the current tenant is a sub-tenant, the new path is created in the specified directory.

       A complete HDFS storage directory can contain a maximum of 1,023 characters. An HDFS directory name contains digits, letters, spaces, and underscores (_). The name cannot start or end with a space.
     - Set **Quota** to the quotas of file and directory quantity.

       Maximum Number of Files/Directories is optional. Its value ranges from **1** to **9223372036854775806**.
     - Set **Storage Space Quota** to the storage space size of the tenant directory.

       The value of **Storage Space Quota** ranges from **1** to **8796093022208**.

       NOTE

       To ensure data reliability, one copy of a file is automatically generated when the file is stored in HDFS. That is, two copies of the same file are stored by default. The HDFS storage space indicates the total disk space occupied by all these copies. For example, if the value of **Storage Space Quota** is set to **500**, the actual space for storing files is about 250 MB (500/2 = 250).
  5. Click **OK**. The system creates tenant directories in the HDFS root directory.
- Modify a tenant directory.
  1. On MRS Manager, click **Tenant**.
  2. In the tenant list on the left, click the tenant whose HDFS storage directory needs to be modified.
  3. Click the **Resource** tab.
  4. In the **HDFS Storage** table, click **Modify** in the **Operation** column of the specified tenant directory.
     - Set **Quota** to the quotas of file and directory quantity.

       Maximum Number of Files/Directories is optional. Its value ranges from **1** to **9223372036854775806**.

- Set **Storage Space Quota** to the storage space size of the tenant directory.

  The value of **Storage Space Quota** ranges from **1** to **8796093022208**.

  NOTE

  To ensure data reliability, one copy of a file is automatically generated when the file is stored in HDFS. That is, two copies of the same file are stored by default. The HDFS storage space indicates the total disk space occupied by all these copies. For example, if the value of **Storage Space Quota** is set to **500**, the actual space for storing files is about 250 MB (500/2 = 250).

e. Click **OK**.

- Delete a tenant directory.

a. On MRS Manager, click **Tenant**.

b. In the tenant list on the left, click the tenant whose HDFS storage directory needs to be deleted.

c. Click the **Resource** tab.

d. In the **HDFS Storage** table, click **Delete** in the **Operation** column of the specified tenant directory.

   The default HDFS storage directory set during tenant creation cannot be deleted. Only the newly added HDFS storage directory can be deleted.

e. Click **OK**.

10.9.6 Restoring Tenant Data

Scenario

Tenant data is stored on Manager and in cluster components by default. When components are restored from faults or reinstalled, some tenant configuration data may be abnormal. In this case, you can manually restore the tenant data.

Procedure

Step 1 On MRS Manager, click **Tenant**.

Step 2 In the tenant list on the left, click a tenant node.

Step 3 Check the status of the tenant data.

1. In **Summary**, check the color of the circle on the left of **Basic Information**. Green indicates that the tenant is available and gray indicates that the tenant is unavailable.
2. Click **Resources** and check the circle on the left of **Yarn** or **HDFS Storage**. Green indicates that the resource is available, and gray indicates that the resource is unavailable.
3. Click **Service Association** and check the **Status** column of the associated service table. **Good** indicates that the component can provide services for the associated tenant. **Bad** indicates that the component cannot provide services for the tenant.

4. If any check result is abnormal, go to **Step 4** to restore tenant data.

Step 4 Click **Restore Tenant Data**.

Step 5 In the **Restore Tenant Data** window, select one or more components whose data needs to be restored. Click **OK**. The system automatically restores the tenant data.

----End

10.9.7 Creating a Resource Pool

Scenario

In an MRS cluster, users can logically divide Yarn cluster nodes to combine multiple NodeManagers into a Yarn resource pool. Each NodeManager belongs to one resource pool only. The system contains a **Default** resource pool by default. All NodeManagers that are not added to customized resource pools belong to this resource pool.

You can create a customized resource pool on MRS Manager and add hosts that have not been added to other customized resource pools to it.

Procedure

Step 1 On MRS Manager, click **Tenant**.

Step 2 Click the **Resource Pools** tab.

Step 3 Click **Add Resource Pool**.

Step 4 In **Create Resource Pool**, set the properties of the resource pool.

- **Name**: Enter a name for the resource pool. The name of the newly created resource pool cannot be **Default**.

  The name consists of 1 to 20 characters and can contain digits, letters, and underscores (_) but cannot start with an underscore (_).

- **Hosts**: In the host list on the left, select the name of a specified host and click >> to add the selected host to the resource pool. Only hosts in the cluster can be selected. The host list of a resource pool can be left blank.

Step 5 Click **OK**.

Step 6 After a resource pool is created, users can view the **Name**, **Members**, **Type**, **vCore** and **Memory** in the resource pool list. Hosts that are added to the customized resource pool are no longer members of the **Default** resource pool.

----End

10.9.8 Modifying a Resource Pool

Scenario

You can modify members of an existing resource pool on MRS Manager.

Procedure

Step 1 On MRS Manager, click **Tenant**.

Step 2 Click the **Resource Pools** tab.

Step 3 Locate the row that contains the specified resource pool, and click **Modify** in the **Operation** column.

Step 4 In **Modify Resource Pool**, modify **Added Hosts**.

- Adding a host: Select the name of a specified host in host list on the left and click >> to add the selected host to the resource pool.
- Deleting a host: In the host list on the right, select the name of a specified host and click << to add the selected host to the resource pool. The host list of a resource pool can be left blank.

Step 5 Click **OK**.

----End

10.9.9 Deleting a Resource Pool

Scenario

You can delete an existing resource pool on MRS Manager.

Prerequisites

- Any queue in a cluster cannot use the resource pool to be deleted as the default resource pool. Before deleting the resource pool, cancel the default resource pool. For details, see **Configuring a Queue**.
- Resource distribution policies of all queues have been cleared from the resource pool being deleted. For details, see **Clearing Configuration of a Queue**.

Procedure

Step 1 On MRS Manager, click **Tenant**.

Step 2 Click the **Resource Pools** tab.

Step 3 Locate the row that contains the specified resource pool, and click **Delete** in the **Operation** column.

In the displayed dialog box, click **OK**.

----End

10.9.10 Configuring a Queue

Scenario

This section describes how to modify the queue configuration for a specified tenant on MRS Manager.

Prerequisites

A tenant associated with Yarn and allocated dynamic resources has been added.

Procedure

Step 1 On MRS Manager, click **Tenant**.

Step 2 Click the **Dynamic Resource Plan** tab.

Step 3 Click the **Queue Configuration** tab.

Step 4 In the tenant queue table, click **Modify** in the **Operation** column of the specified tenant queue.

NOTE

In the tenant list on the left of the **Tenant Management** tab, click the target tenant. In the window that is displayed, choose **Resource**. On the page that is displayed, click ✎ to open the queue modification page.

Table 10-32 Queue configuration parameters

Parameter	Description
Maximum Application	Specifies the maximum number of applications. The value ranges from 1 to 2147483647.
Maximum AM Resource Percent	Specifies the maximum percentage of resources that can be used to run the ApplicationMaster in a cluster. The value ranges from 0 to 1.
Minimum User Limit Percent (%)	Specifies the minimum percentage of resources consumed by a user. The value ranges from 0 to 100.
User Limit Factor	Specifies the limit factor of the maximum user resource usage. The maximum user resource usage percentage can be obtained by multiplying the limit factor with the percentage of the tenant's actual resource usage in the cluster. The minimum value is **0**.
Status	Specifies the current status of a resource plan. The values are **Running** and **Stopped**.

Parameter	Description
Default Resource Pool	Specifies the resource pool used by a queue. The default value is **Default**. If you want to change the resource pool, configure the queue capacity first. For details, see **Configuring the Queue Capacity Policy of a Resource Pool**.

----End

10.9.11 Configuring the Queue Capacity Policy of a Resource Pool

Scenario

After a resource pool is added, the capacity policies of available resources need to be configured for Yarn task queues. This ensures that tasks in the resource pool are running properly. Each queue can be configured with the queue capacity policy of only one resource pool. Users can view the queues in any resource pool and configure queue capacity policies. After the queue policies are configured, Yarn task queues and resource pools are associated.

You can configure queue policies on MRS Manager.

Prerequisites

- A resource pool has been added.
- The task queues are not associated with other resource pools. By default, all queues are associated with the **Default** resource pool.

Procedure

Step 1 On MRS Manager, click **Tenant**.

Step 2 Click the **Dynamic Resource Plan** tab.

Step 3 In **Resource Pools**, select a specified resource pool.

Available Resource Quota: indicates that all resources in each resource pool are available for queues by default.

Step 4 Locate the specified queue in the **Resource Allocation** table, and click **Modify** in the **Operation** column.

Step 5 In **Modify Resource Allocation**, configure the resource capacity policy of the task queue in the resource pool.

- **Capacity (%)**: specifies the percentage of the current tenant's computing resource usage.
- **Maximum Capacity (%)**: specifies the percentage of the current tenant's maximum computing resource usage.

Step 6 Click **OK** to save the settings.

----End

10.9.12 Clearing Configuration of a Queue

Scenario

Users can clear the configuration of a queue on MRS Manager when the queue does not need resources from a resource pool or if a resource pool needs to be disassociated from the queue. Clearing queue configurations means that the resource capacity policy of the queue is canceled.

Prerequisites

If a queue is to be unbound from a resource pool, this resource pool cannot serve as the default resource pool of the queue. Therefore, you must first change the default resource pool of the queue to another one. For details, see **Configuring a Queue**.

Procedure

Step 1 On MRS Manager, click **Tenant**.

Step 2 Click the **Dynamic Resource Plan** tab.

Step 3 In **Resource Pools**, select a specified resource pool.

Step 4 Locate the specified queue in the **Resource Allocation** table, and click **Clear** in the **Operation** column.

In the **Clear Queue Configuration** dialog box, click **OK** to clear the queue configuration in the current resource pool.

NOTE

If no resource capacity policy is configured for a queue, the clearing function is unavailable for the queue by default.

----End

10.10 Backup and Restoration

10.10.1 Introduction

Purpose

MRS Manager provides backup and restoration for user data and system data. The backup function is provided based on components to back up Manager data (including OMS data and LdapServer data), Hive user data, component metadata saved in DBService, and HDFS metadata.

Backup and restoration tasks are performed in the following scenarios:

- Routine backup is performed to ensure the data security of the system and components.
- If the system is faulty, the data backup can be used to recover the system.
- If the active cluster is completely faulty, a mirror cluster identical to the active cluster needs to be created. You can use the backup data to restore the active cluster.

Table 10-33 Backing up metadata

Backup Type	Backup Content
OMS	Database data (excluding alarm data) and configuration data in the cluster management system to be backed up by default
LdapServer	User information, including the username, password, key, password policy, and group information
DBService	Metadata of the components (Hive) managed by DBService
NameNode	HDFS metadata.

Principles

Task

Before backup or restoration, you need to create a backup or restoration task and set task parameters, such as the task name, backup data source, and type of backup file save path. Data backup and restoration can be performed by executing backup and restoration tasks. When the Manager is used to recover the data of HDFS, HBase, Hive, and NameNode, no cluster can be accessed.

Each backup task can back up data of different data sources and generates an independent backup file for each data source. All the backup files generated in each backup task form a backup file set, which can be used in restoration tasks. Backup data can be stored on Linux local disks, local cluster HDFS, and standby cluster HDFS. The backup task provides the full backup or incremental backup policies. HDFS and Hive backup tasks support the incremental backup policy, while OMS, LdapServer, DBService, and NameNode backup tasks support only the full backup policy.

NOTE

Task execution rules:

- If a task is being executed, the task cannot be executed repeatedly and other tasks cannot be started at the same time.
- The interval at which a periodical task is automatically executed must be greater than 120s; otherwise, the task is postponed and will be executed in the next period. Manual tasks can be executed at any interval.
- When a periodic task is to be automatically executed, the current time cannot be 120s later than the task start time; otherwise, the task is postponed and executed in the next period.
- When a periodic task is locked, it cannot be automatically executed and needs to be manually unlocked.
- Before an OMS, LdapServer, DBService, or NameNode backup task starts, ensure that the LocalBackup partition on the active management node has more than 20 GB available space. Otherwise, the backup task cannot be started.
- When you are planning backup and restoration tasks, select the data to be backed up or restored strictly based on the service logic, data store structure, and database or table association. The system creates a default periodic backup task **default** whose execution interval is 24 hours to perform full backup of OMS, LdapServer, DBService, and NameNode data to the Linux local disk.

Specifications

Table 10-34 Backup and restoration feature specifications

Item	Specifications
Maximum number of backup or restoration tasks	100
Number of concurrent running tasks	1
Maximum number of waiting tasks	199
Maximum size of backup files on a Linux local disk (GB)	600

Table 10-35 Specifications of the **default** task

Item	OMS	LdapServer	DBService	NameNode
Backup period	1 hour			
Maximum number of copies	2			
Maximum size of a backup file	10 MB	20 MB	100 MB	1.5 GB

Item	OMS	LdapServer	DBService	NameNode
Maximum size of disk space used	20 MB	40 MB	200 MB	3 GB
Save path of backup data	*Data save path*/**LocalBackup/** of the active and standby management nodes			

NOTE

The backup data of the **default** task must be periodically transferred and saved outside the cluster based on the enterprise O&M requirements.

10.10.2 Backing Up Metadata

Scenario

To ensure the security of metadata either on a routine basis or before and after performing critical metadata operations (such as scale-out, scale-in, patch installation, upgrades, and migration), metadata must be backed up. The backup data can be used to recover the system if an exception occurs or if the operation has not achieved the expected result. This minimizes the adverse impact on services. Metadata includes data of OMS, LdapServer, DBService, and NameNode. MRS Manager data to be backed up includes OMS data and LdapServer data.

By default, metadata backup is supported by the **default** task. This section describes how to create a backup task and back up metadata on MRS Manager. Both automatic backup tasks and manual backup tasks are supported.

Prerequisites

- A standby cluster for backing up data has been created, and the network is connected. The inbound rules of the two security groups on the peer cluster have been added to the two security groups in each cluster to allow all access requests of all protocols and ports of all ECSs in the security groups.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path*/**LocalBackup/** has sufficient space on the active and standby management nodes.

Procedure

Step 1 Create a backup task.

1. On MRS Manager, choose **System** > **Back Up Data**.
2. Click **Create Backup Task**.

Step 2 Configure a backup policy.

1. Set **Task Name** to the name of the backup task.

2. Set **Backup Mode** to the type of the backup task. **Periodic** indicates that the backup task is periodically executed. **Manual** indicates that the backup task is manually executed.

   To create a periodic backup task, set the following parameters:

   – **Started**: indicates the time when the task is started for the first time.

   – **Period**: indicates the task execution interval. The options include **By hour** and **By day**.

   – **Backup Policy**: indicates the volume of data to be backed up in each task execution. The options include **Full backup at the first time and incremental backup later**, **Full backup every time**, and **Full backup once every n times**. If you select **Full backup once every n times**, you need to specify the value of **n**.

Step 3 Select backup sources.

In the **Configuration** area, select **OMS** and **LdapServer** under **Metadata**.

Step 4 Set backup parameters.

1. Set **Path Type** of **OMS** and **LdapServer** to a backup directory type.

   The following backup directory types are supported:

   – **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files. By default, the backup files are stored in *Data storage path*/**LocalBackup/**. If you select **LocalDir**, you need to set the maximum number of copies to specify the number of backup files that can be retained in the backup directory.

   – **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster. If you select **SFTP**, set the following parameters:

     ▪ **Target Path**: indicates the HDFS directory for storing the backup files. The save path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory.

     ▪ **Max Number of Backup Copies**: indicates the number of backup files that can be retained in the backup directory.

     ▪ **Target Instance Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

2. Click **OK**.

Step 5 Execute the backup task.

In the **Operation** column of the created task in the backup task list, click **Back Up Now** if **Backup Mode** is set to **Periodic** or click **Start** If **Backup Mode** is set to **Manual** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files. The format of the backup file name is *Version_Data source_Task execution time***.tar.gz**.

----End

10.10.3 Restoring Metadata

Scenario

You need to restore metadata in the following scenarios: A user modifies or deletes data unexpectedly, data needs to be retrieved, system data becomes abnormal or does not achieve the expected result, all modules are faulty, and data is migrated to a new cluster.

This section describes how to restore metadata on MRS Manager. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To restore the data when services are normal, manually back up the latest management data first and then restore the data. Otherwise, the data that is generated after the data backup and before the data restoration will be lost.
- Use the OMS data and LdapServer data backed up at the same time to restore data. Otherwise, the service and operation may fail.
- By default, MRS clusters use DBService to store Hive metadata.

Impact on the System

- After the data is restored, the data generated between the backup time and restoration time is lost.
- After the data is restored, the configuration of the components that depend on DBService may expire and these components need to be restarted.

Prerequisites

- The data in the OMS and LdapServer backup files has been backed up at the same time.
- The status of the OMS resources and the LdapServer instances is normal. If the status is abnormal, data restoration cannot be performed.
- The status of the cluster hosts and services is normal. If the status is abnormal, data restoration cannot be performed.
- The cluster host topologies during data restoration and data backup are the same. If the topologies are different, data restoration cannot be performed and you need to back up data again.
- The services added to the cluster during data restoration and data backup are the same. If the services are different, data restoration cannot be performed and you need to back up data again.
- The status of the active and standby DBService instances is normal. If the status is abnormal, data restoration cannot be performed.
- The upper-layer applications depending on the MRS cluster have been stopped.

- On MRS Manager, you have stopped all the NameNode role instances whose data is to be recovered. Other HDFS role instances are running properly. After data is recovered, the NameNode role instances need to be restarted and cannot be accessed before the restart.
- You have checked whether NameNode backup files have been stored in the *Data save path***/LocalBackup/** directory on the active management node.

Procedure

Step 1 Check the location of backup data.

1. On MRS Manager, choose **System** > **Back Up Data**.
2. In the row where the specified backup task resides, choose **More** > **View History** in the **Operation** column to display the historical execution records of the backup task. In the window that is displayed, select a success record and click **View Backup Path** in the corresponding column to view its backup path information. Find the following information:
   - **Backup Object**: indicates the backup data source.
   - **Backup Path**: indicates the full path where backup files are stored.
3. Select the correct path, and manually copy the full path of backup files in **Backup Path**.

Step 2 Create a restoration task.

1. On MRS Manager, choose System > Recovery Management.
2. On the page that is displayed, click **Create Restoration Task**.
3. Set **Task Name** to the name of the restoration task.

Step 3 Select restoration sources.

In **Configuration**, select the metadata component whose data is to be restored.

Step 4 Set the restoration parameters.

1. Set **Path Type** to a backup directory type.
2. The settings vary according to backup directory types:
   - **LocalDir**: indicates that the backup files are stored on the local disk of the active management node. If you select **LocalDir**, you need to set **Source Path** to specify the full path of the backup file. For example, *Data storage path***/LocalBackup/***Backup task name_Task creation time***/***Data source_Task execution time***/***Version number_Data source_Task execution time***.tar.gz**.
   - **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster. If you select **SFTP**, set the following parameters:
     - **Source Path**: indicates the full HDFS path of a backup file. for example, *Backup path/Backup task name_Task creation time/Version_Data source_Task execution time***.tar.gz**.
     - **Source Instance Name**: indicates the name of NameService corresponding to the backup directory when a restoration task is being executed. The default value is **hacluster**.

3. Click **OK**.

Step 5 Execute the restoration task.

In the restoration task list, locate the row where the created task resides, and click **Start** in the **Operation** column.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and try to execute the task again by clicking **Start**.

Step 6 Determine what metadata has been restored.

- If the OMS and LdapServer metadata is restored, go to **Step 7**.
- If DBService data is restored, no further action is required.
- Restore NameNode data. On MRS Manager, choose **Services > HDFS > More > Restart Service**. The task is complete.

Step 7 Restarting Manager for the recovered data to take effect

1. In MRS Manager, Choose **LdapServer** > **More** > **Restart Service** and click **OK**. Wait until the LdapServer service is restarted successfully.
2. Log in to the active management node. For details, see **Determining Active and Standby Management Nodes of Manager**.
3. Run the following command to restart OMS:

   sh ${BIGDATA_HOME}/om-0.0.1/sbin/restart-oms.sh

   The command has been executed successfully if the following information is displayed:

   ```
   start HA successfully.
   ```
4. On MRS Manager, choose **KrbServer > More > Synchronize Configuration**. Do not select Restart the services and instances whose configuration has expired. Click **OK** and wait until the KrbServer service configuration is synchronized and restarted successfully.
5. Choose **Services > More > Synchronize Configuration**. Do not select Restart the services and instances whose configuration has expired. Click **OK** and wait until the cluster is configured and synchronized successfully.
6. Choose **Services > More > Stop Cluster**. After the cluster is stopped, choose **Services > More> Start Cluster**.

----End

10.10.4 Modifying a Backup Task

Scenario

This section describes how to modify the parameters of a created backup task on MRS Manager to meet changing service requirements. The parameters of restoration tasks can be viewed but not modified.

Impact on the System

After a backup task is modified, the new parameters take effect when the task is executed next time.

Prerequisites

- A backup task has been created.
- A new backup task policy has been planned based on the actual situation.

Procedure

Step 1 On MRS Manager, choose **System** > **Back Up Data**.

Step 2 In the task list, locate a specified task, click **Modify** in the **Operation** column to go to the configuration modification page.

Step 3 Modify the following parameters on the displayed page:

- Manual backup:
  - Target Path
  - Max Number of Backup Copies
- Periodic backup:
  - Started
  - Period
  - Target Path
  - Max Number of Backup Copies

NOTE

- When **Path Type** is set to **LocalHDFS**, **Target Path** is valid for modifying a backup task.
- After you change the value of **Target Path** for a backup task, full backup is performed by default when the task is executed for the first time.

Step 4 Click **OK**.

----End

10.10.5 Viewing Backup and Restoration Tasks

Scenario

This section describes how to view created backup and restoration tasks and check their running status on MRS Manager.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 Click **Back Up Data** or **Restore Data**.

Step 3 In the task list, obtain the previous execution result in the **Task Progress** column. Green indicates that the task is executed successfully, and red indicates that the execution fails.

Step 4 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view the historical record of backup and restoration execution.

In the displayed window, click **View** in the **Details** column. The task execution logs and paths are displayed.

----End

Related Tasks

- Modifying a backup task

  For details, see **Modifying a Backup Task**.
- Viewing a restoration task

  In the **Operation** column of the specified task in the task list, click **View Details** to view the restoration task. You can only view but cannot modify the parameters of a restoration task.
- Executing a backup or restoration task

  In the task list, locate a specified task and click **Start** in the **Operation** column to start a backup or restoration task that is ready or fails to be executed. Executed restoration tasks cannot be repeatedly executed.
- Stopping backup tasks

  In the task list, locate a specified task and click **More** > **Stop** in the **Operation** column to stop a backup task that is running.
- Deleting a backup or restoration task

  In the **Operation** column of the specified task in the task list, choose **More** > **Delete** to delete the backup or restoration task. After a task is deleted, the backup data is retained by default.
- Suspending a backup task

  In the **Operation** column of the specified task in the task list, choose **More** > **Suspend** to suspend the backup task. Only periodic backup tasks can be suspended. Suspended backup tasks are no longer executed automatically. When you suspend a backup task that is being executed, the task execution stops. To cancel the suspension status of a task, click **More** > **Resume**.

10.11 Security Management

10.11.1 Default Users of Clusters with Kerberos Authentication Disabled

User Classification

The MRS cluster provides the following two types of users. Users are advised to periodically change the passwords. It is not recommended to use the default passwords.

User Type	Description
System users	User who runs OMS processes
Database users	● User who manages OMS database and accesses data ● User who runs the database of service components (Hive, Loader, and DBService)

System users

NOTE

- User **ldap** of the OS is required in the MRS cluster. Do not delete this account. Otherwise, the cluster may not work properly. Password management policies are maintained by the operation users.
- Reset the passwords when you change the passwords of user **ommdba** and user **omm** for the first time. Change the passwords periodically after retrieving them.

Operation	Username	Initial Password	Description
MRS cluster node OS user	root	Set by the user	User for logging in to the node in the MRS cluster. This user is an OS user generated on all nodes.

User Group Information

Default User Group	Description
supergroup	Primary group of user **admin**, which has no additional permissions in the cluster with Kerberos authentication disabled.
check_sec_ldap	Used to test whether the active LDAP works properly. This user group is generated randomly in a test and automatically deleted after the test is complete. which is an internal system user group used only between components.
Manager_tenant	Tenant system user group, which is an internal system user group used only between components. It is used only in clusters with Kerberos authentication enabled.
System_administrator	MRS cluster system administrator group, which is an internal system user group used only between components. It is used only in clusters with Kerberos authentication enabled.

Default User Group	Description
Manager_viewer	MRS Manager system viewer group, which is an internal system user group used only between components. It is used only in clusters with Kerberos authentication enabled.
Manager_operator	MRS Manager system operator group, which is an internal system user group used only between components. It is used only in clusters with Kerberos authentication enabled.
Manager_auditor	MRS Manager system auditor group, which is an internal system user group used only between components. It is used only in clusters with Kerberos authentication enabled.
Manager_administrator	MRS Manager system administrator group, which is an internal system user group used only between components. It is used only in clusters with Kerberos authentication enabled.
compcommon	MRS cluster internal group, used to access public resources in the cluster. All system users and system running users are added to this user group by default.
default_1000	User group created for tenants, which is an internal system user group used only between components.
launcher-job	MRS internal group, which is used to submit jobs using V2 APIs.

OS User Group	Description
wheel	Primary group of MRS internal running user **omm**.
ficommon	MRS cluster common group that corresponds to **compcommon** for accessing public resource files stored in the OS of the cluster.

Database users

MRS cluster system database users include OMS database users and DBService database users.

NOTE

Do not delete database users. Otherwise, the cluster or components may not work properly.

Operation	Default User	Initial Password	Description
OMS database	ommdba	dbChangeMe@123456	OMS database administrator who performs maintenance operations, such as creating, starting, and stopping.
	omm	ChangeMe@123456	User for accessing OMS database data
DBService database	omm	dbserverAdmin@123	Administrator of the GaussDB database in the DBService component
	hive	HiveUser@	User for Hive to connect to the DBService database
	hue	HueUser@123	User for Hue to connect to the DBService database

10.11.2 Default Users of Clusters with Kerberos Authentication Enabled

User Classification

The MRS cluster provides the following three types of users. Users are advised to periodically change the passwords. It is not recommended to use the default passwords.

User Type	Description
System user	● User created on Manager for MRS cluster O&M and service scenarios. There are two types of users: – **Human-machine** user: used for Manager O&M scenarios and component client operation scenarios. – **Machine-machine** user: used for MRS cluster application development scenarios. ● User who runs OMS processes.
Internal system user	Internal user who performs process communications, saves user group information, and associates user permissions.
Database user	● User who manages OMS database and accesses data. ● User who runs the database of service components (Hive, Hue, Loader, and DBService)

System User

NOTE

- User **ldap** of the OS is required in the MRS cluster. Do not delete this account. Otherwise, the cluster may not work properly. Password management policies are maintained by the operation users.
- Reset the passwords when you change the passwords of user **ommdba** and user **omm** for the first time. Change the passwords periodically after retrieving them.

Type	Username	Initial Password	Description
System administrator of the MRS cluster	admin	Specified by the user during the cluster creation.	Manager administrator with the following permissions: ● Common HDFS and ZooKeeper user permissions. ● Permissions to submit and query MapReduce and Yarn tasks, manage Yarn queues, and access the Yarn web UI. ● Permissions to submit, query, activate, deactivate, reassign, delete topologies, and operate all topologies of the Storm service. ● Permissions to create, delete, authorize, reassign, consume, write, and query topics of the Kafka service.
MRS cluster node OS user	omm	Randomly generated by the system.	Internal running user of the MRS cluster system. This user is an OS user generated on all nodes and does not require a unified password.
MRS cluster node OS user	root	Set by the user.	User for logging in to the node in the MRS cluster. This user is an OS user generated on all nodes.

Internal System Users

NOTE

Do not delete the following internal system users. Otherwise, the cluster or components may not work properly.

Type	Default User	Initial Password	Description
Component running user	hdfs	Hdfs@123	This user is the HDFS system administrator and has the following permissions: 1. File system operation permissions: ● Views, modifies, and creates files. ● Views and creates directories. ● Views and modifies the groups where files belong. ● Views and sets disk quotas for users. 2. HDFS management operation permissions: ● Views the web UI status. ● Views and sets the active and standby HDFS status. ● Enters and exits the HDFS in security mode. ● Checks the HDFS file system.

Type	Default User	Initial Password	Description
	hbase	Hbase@123	This user is the HBase system administrator and has the following permissions: ● Cluster management permission: **Enable** and **Disable** operations on tables to trigger MajorCompact and ACL operations. ● Grants and revokes permissions, and shuts down the cluster. ● Table management permission: Creates, modifies, and deletes tables. ● Data management permission: Reads and writes data in tables, column families, and columns. ● Accesses the HBase web UI.
	mapred	Mapred@123	This user is the MapReduce system administrator and has the following permissions: ● Submits, stops, and views the MapReduce tasks. ● Modifies the Yarn configuration parameters. ● Accesses the Yarn and MapReduce web UI.
	spark	Spark@123	This user is the Spark system administrator and has the following permissions: ● Accesses the Spark web UI. ● Submits Spark tasks.

User Group Information

Default User Group	Description
hadoop	Users added to this user group have the permission to submit tasks to all Yarn queues.
hbase	Common user group. Users added to this user group will not have any additional permission.
hive	Users added to this user group can use Hive.
spark	Common user group. Users added to this user group will not have any additional permission.
supergroup	Users added to this user group can have the administrator permission of HBase, HDFS, and Yarn and can use Hive.
check_sec_ldap	Used to test whether the active LDAP works properly. This user group is generated randomly in a test and automatically deleted after the test is complete. This is an internal system user group used only between components.
Manager_tenant	Tenant system user group, which is an internal system user group used only between components.
System_administrator	MRS cluster system administrator group, which is an internal system user group used only between components.
Manager_viewer	MRS Manager system viewer group, which is an internal system user group used only between components.
Manager_operator	MRS Manager system operator group, which is an internal system user group used only between components.
Manager_auditor	MRS Manager system auditor group, which is an internal system user group used only between components.
Manager_administrator	MRS Manager system administrator group, which is an internal system user group used only between components.
compcommon	Internal system group for accessing public resources in a cluster. All system users and system running users are added to this user group by default.
default_1000	User group created for tenants, which is an internal system user group used only between components.

Default User Group	Description
kafka	Kafka common user group. Users added to this group need to be granted with read and write permission by users in the **kafkaadmin** group before accessing the desired topics.
kafkasuperuser	Users added to this group have permissions to read data from and write data to all topics.
kafkaadmin	Kafka administrator group. Users added to this group have the permissions to create, delete, authorize, as well as read from and write data to all topics.
storm	Storm common user group. Users added to this group have the permissions to submit topologies and manage their own topologies.
stormadmin	Storm administrator user group. Users added to this group have the permissions to submit topologies and manage their own topologies.
opentsdb	Common user group. Users added to this user group will not have any additional permission.
presto	Common user group. Users added to this user group will not have any additional permission.
flume	Common user group. Users added to this user group will not have any additional permission.
launcher-job	MRS internal group, which is used to submit jobs using V2 APIs.

OS User Group	Description
wheel	Primary group of MRS internal running user **omm**.
ficommon	MRS cluster common group that corresponds to **compcommon** for accessing public resource files stored in the OS of the cluster.

Database User

MRS cluster system database users include OMS database users and DBService database users.

NOTE

Do not delete database users. Otherwise, the cluster or components may not work properly.

Type	Default User	Initial Password	Description
OMS database	ommdba	dbChangeMe@123456	OMS database administrator who performs maintenance operations, such as creating, starting, and stopping applications.
	omm	ChangeMe@123456	User for accessing OMS database data.
DBService database	omm	dbserverAdmin@123	Administrator of the GaussDB database in the DBService component.
	hive	HiveUser@	User for Hive to connect to the DBService database.
	hue	HueUser@123	User for Hue to connect to the DBService database.
	sqoop	SqoopUser@	User for Loader to connect to the DBService database.
	ranger	RangerUser@	User for Ranger to connect to the DBService database.

10.11.3 Changing the Password of an OS User

Scenario

This section describes how to periodically change the login passwords of the OS users **omm**, **ommdba**, and **root** on MRS cluster nodes to improve the system O&M security.

Passwords of users **omm**, **ommdba**, and **root** on each node can be different.

Procedure

Step 1 Log in to the **Master1** node and then log in to other nodes whose OS user passwords need to be changed.

Step 2 Run the following command to switch to user **root**:

sudo su - root

Step 3 Run the following command to change the passwords of users **omm**, **ommdba**, or **root**:

passwd omm

passwd ommdba

passwd root

For example, if you run the **omm:passwd** command, the system displays the following information:

```
Changing password for user omm.
New password:
```

Enter a new password. The password change policies for an OS vary according to the OS that is used.

```
Retype new password:
passwd: all authentication tokens updated successfully.
```

NOTE

The default password complexity requirements of the MRS cluster are as follows:

- The password must contain at least eight characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, spaces, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The new password cannot be the same as last five historical passwords.

----End

10.11.4 Changing the password of user admin

This section describes how to periodically change the password of cluster user **admin** to improve the system O&M security.

If the password is changed, the downloaded user credential will be unavailable. Download the authentication credential again, and replace the old one.

Changing the Password of User admin on the Cluster Node

Step 1 Update the client of the active management node. For details, see **Updating a Client (Versions Earlier Than 3.x)**.

Step 2 Log in to the active management node.

Step 3 (Optional) To change the password as user **omm**, run the following command to switch the user:

sudo su - omm

Step 4 Run the following command to switch to the client directory, for example, **/opt/client**.

cd /opt/client

Step 5 Run the following command to configure environment variables:

source bigdata_env

Step 6 Run the following command to change the password of user **admin**: This operation takes effect in the whole cluster.

kpasswd admin

Enter the old password and then enter a new password twice.

For the cluster, the default password complexity requirements are as follows:

- The password must contain at least eight characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, spaces, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The password cannot be the username or the reverse username.

----End

Changing the Password of User admin on MRS Manager

You can change the password of user **admin** on MRS Manager only for clusters with Kerberos authentication enabled and clusters with Kerberos authentication disabled but the EIP function enabled.

Step 1 Log in to MRS Manager as user **admin**.

Step 2 Click the username in the upper right corner of the page and choose **Change Password**.

Step 3 On the **Change Password** page, set **Old Password**, **New Password**, and **Confirm Password**.

NOTE

The default password complexity requirements are as follows:

- The password must contain 8 to 32 characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, spaces, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The password cannot be the username or the reverse username.

Step 4 Click **OK**. Log in to MRS Manager with the new password.

----End

Resetting the Password for User admin

Step 1 Log in to the **Master1** node.

Step 2 (Optional) To change the password as user **omm**, run the following command to switch the user:

sudo su - omm

Step 3 Run the following command to switch to the client directory, for example, **/opt/client**:

cd /opt/client

Step 4 Run the following command to configure environment variables:

source bigdata_env

Step 5 Run the following command to log in to the console as user **kadmin/admin**:

kadmin -p kadmin/admin

NOTE

The default password of user **kadmin/admin** is **KAdmin@123**, which will expire upon your first login. Change the password as prompted and keep the new password secure.

Step 6 Run the following command to reset the password of a component running user. This operation takes effect for all servers.

cpw *Component running user name*

For example, to reset the password of user admin, run the **cpw admin** command.

For the cluster, the default password complexity requirements are as follows:

- The password must contain 8 to 32 characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, spaces, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The password cannot be the username or the reverse username.

----End

10.11.5 Changing the Password of the Kerberos Administrator

Scenario

This section describes how to periodically change the password of the Kerberos administrator **kadmin** of the MRS cluster to improve the system O&M security.

If the password is changed, the downloaded user credential will be unavailable. Download the authentication credential again, and replace the old one.

Prerequisites

A client has been prepared on the **Master1** node.

Procedure

Step 1 Log in to the **Master1** node.

Step 2 (Optional) To change the password as user **omm**, run the following command to switch the user:

sudo su - omm

Step 3 Run the following command to switch to the client directory, for example, **/opt/client**.

cd /opt/client

Step 4 Run the following command to configure environment variables:

source bigdata_env

Step 5 Run the following command to change the password of **kadmin/admin**. This operation takes effect for all servers.

kpasswd kadmin/admin

For the cluster, the default password complexity requirements are as follows:

- The password must contain at least eight characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, spaces, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The password cannot be the username or the reverse username.

----End

10.11.6 Changing the Passwords of the LDAP Administrator and the LDAP User

Scenario

This section describes how to periodically change the passwords of the LDAP administrator **rootdn:cn=root,dc=hadoop,dc=com** and the LDAP user **pg_search_dn:cn=pg_search_dn,ou=Users,dc=hadoop,dc=com** to improve the system O&M security.

Impact on the System

All services need to be restarted for the new password to take effect. The services are unavailable during the restart.

Procedure

Step 1 On MRS Manager, choose **Services > LdapServer > More**.

Step 2 Click **Change Password**.

Step 3 In the **Change Password** dialog box, select the user whose password needs to be modified in the **User Information** drop-down box.

Step 4 Enter the old password in the **Old Password** text box, and enter the new password in the **New Password** and **Confirm Password** text boxes.

The default password complexity requirements are as follows:

- The password contains 16 to 32 characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, and special characters (`~!@#$%^&*()-_=+\|[{}];:",<.>/?).
- The password cannot be the username or the reverse username.
- The new password cannot be the same as the current password.

NOTE

The default password of the LDAP administrator **rootdn:cn=root,dc=hadoop,dc=com** is **LdapChangeMe@123**, and that of the LDAP user **pg_search_dn:cn=pg_search_dn,ou=Users,dc=hadoop,dc=com** is **pg_search_dn@123**. Periodically change the passwords and keep them secure.

Step 5 Select **I have read the information and understand the impact**, and click **OK** to confirm the modification and restart the service.

----End

10.11.7 Changing the Password of a Component Running User

Scenario

This section describes how to periodically change the password of the component running user of the MRS cluster to improve the system O&M security.

If the initial password is randomly generated by the system, reset the password.

If the password is changed, the downloaded user credential will be unavailable. Download the authentication credential again, and replace the old one.

Prerequisites

A client has been prepared on the **Master1** node.

Procedure

Step 1 Log in to the **Master1** node.

Step 2 (Optional) To change the password as user **omm**, run the following command to switch the user:

sudo su - omm

Step 3 Run the following command to switch to the client directory, for example, **/opt/client**:

cd /opt/client

Step 4 Run the following command to configure environment variables:

source bigdata_env

Step 5 Run the following command to log in to the console as user **kadmin/admin**:

kadmin -p kadmin/admin

NOTE

The default password of user **kadmin/admin** is **KAdmin@123**, which will expire upon your first login. Change the password as prompted and keep the new password secure.

Step 6 Run the following command to reset the password of a component running user. This operation takes effect for all servers.

cpw *Component running user name*

For example, to reset the password of user admin, run the **cpw admin** command.

For the cluster, the default password complexity requirements are as follows:

- The password must contain 8 to 32 characters.

- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, spaces, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The password cannot be the username or the reverse username.

----End

10.11.8 Changing the Password of the OMS Database Administrator

Scenario

This section describes how to periodically change the password of the OMS database administrator to improve the system O&M security.

Procedure

Step 1 Log in to the active management node.

NOTE

The password of user **ommdba** cannot be changed on the standby management node. Otherwise, the cluster may not work properly. Change the password on the active management node only.

Step 2 Run the following command to switch the user:

sudo su - omm

Step 3 Run the following command to switch the directory:

cd $OMS_RUN_PATH/tools

Step 4 Run the following command to change the password of user **ommdba**:

mod_db_passwd ommdba

Step 5 Enter the old password of user **ommdba** and enter a new password twice.

The password complexity requirements are as follows:

- The password contains 16 to 32 characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The password cannot be the username or the reverse username.
- The password cannot be the same as the last 20 historical passwords.

If the following information is displayed, the password is changed successfully.

```
Congratulations, update [ommdba] password successfully.
```

----End

10.11.9 Changing the Password of the Data Access User of the OMS Database

Scenario

This section describes how to periodically change the password of the data access user of the OMS database to improve the system O&M security.

Impact on the System

The OMS service needs to be restarted for the new password to take effect. The service is unavailable during the restart.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Change OMS Database Password**.

Step 3 Locate the row that contains user **omm**, and click **Change password** in the **Operation** column.

The password complexity requirements are as follows:

- The password must contain 8 to 32 characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The password cannot be the username or the reverse username.
- The password cannot be the same as the last 20 historical passwords.

Step 4 Click **OK**. When **Operation successful** is displayed, click **Finish**.

Step 5 Locate the row that contains user **omm**, and click **Restart the OMS service** in the **Operation** column to restart the OMS database.

NOTE

If the password is changed but the OMS database is not restarted, the status of user **omm** changes to **Waiting to restart** and the password cannot be changed until the OMS database is restarted.

Step 6 In the displayed dialog box, select **I have read the information and understand the impact**. Click **OK**, and restart the OMS service.

----End

10.11.10 Changing the Password of a Component Database User

Scenario

This section describes how to periodically change the password of the component database user to improve the system O&M security.

Impact on the System

The services need to be restarted for the new password to take effect. The services are unavailable during the restart.

Procedure

Step 1 On MRS Manager, click **Services** and click the name of the database user service to be modified.

Step 2 Determine the component database user whose password is to be changed.

- To change the password of the DBService database user, go to **Step 3**.
- To change the password of the Loader, Hive, or Hue database user, stop the service first and then execute **Step 3**.

Click **Stop Service**.

Step 3 Choose **More** > **Change Password**.

Step 4 Enter the old and new passwords as prompted.

The password complexity requirements are as follows:

- The password of the DBService database user contains 16 to 32 characters. The password of the Loader, Hive, or Hue database user contains 8 to 32 characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The password cannot be the username or the reverse username.
- The password cannot be the same as the last 20 historical passwords.

Step 5 Click **OK**. The system automatically restarts the corresponding service. When **Operation successful** is displayed, click **Finish**.

----End

10.11.11 Replacing the HA Certificate

Scenario

HA certificates are used to encrypt the communication between active/standby processes and HA processes to ensure the communication security. This section describes how to replace the HA certificates on the active and standby management nodes on MRS Manager to ensure the product security.

The certificate file and key file can be generated by the user.

Impact on the System

MRS Manager needs to be restarted during the replacement and cannot be accessed or provide services at that time.

Prerequisites

- You have obtained the **root-ca.crt** HA root certificate file and the **root-ca.pem** key file to be replaced.
- You have prepared a password, such as **Userpwd@123**, for accessing the key file.

  To avoid potential security risks, the password must meet the following complexity requirements:

  – The password must contain at least eight characters.

  – The password must contain at least four types of the following characters: uppercase letters, lowercase letters, digits, and special characters (~`!?,.:;-_'(){}[]/<>@#$%^&*+|\=).

Procedure

Step 1 Log in to the active management node.

Step 2 Run the following commands to switch the user:

sudo su - root

su - omm

Step 3 Run the following commands to generate **root-ca.crt** and **root-ca.pem** in the **${OMS_RUN_PATH}/workspace0/ha/local/cert** directory on the active management node:

sh ${OMS_RUN_PATH}/workspace/ha/module/hacom/script/gen-cert.sh --root-ca --country=*country* **--state=***state* **--city=***city* **--company=***company* **--organize=***organize* **--common-name=***commonname* **--email=***Administrator email address* **--password=***password*

For example, run the **sh ${OMS_RUN_PATH}/workspace/ha/module/hacom/script/gen-cert.sh --root-ca --country=XX--state=XX --city=XX --company=XX --organize=IT --common-name=HADOOP.COM --email=abc@XXX.com --password=Userpwd@123** command.

The command has been executed successfully if the following information is displayed:

```
Generate root-ca pair success.
```

Step 4 On the active management node, run the following command as user **omm** to copy **root-ca.crt** and **root-ca.pem** to the **${BIGDATA_HOME}/om-0.0.1/security/certHA** directory:

cp -arp ${OMS_RUN_PATH}/workspace0/ha/local/cert/root-ca.* ${BIGDATA_HOME}/om-0.0.1/security/certHA

Step 5 Copy **root-ca.crt** and **root-ca.pem** generated on the active management node to the **${BIGDATA_HOME}/om-0.0.1/security/certHA** directory on the standby management node as user **omm**.

Step 6 Run the following command to generate an HA certificate and perform the automatic replacement:

sh ${BIGDATA_HOME}/om-0.0.1/sbin/replacehaSSLCert.sh

Enter the password as prompted, and press **Enter**.

```
Please input ha ssl cert password:
```

The HA certificate is replaced successfully if the following information is displayed:

```
[INFO] Succeed to replace ha ssl cert.
```

Step 7 Run the following command to restart OMS:

sh ${BIGDATA_HOME}/om-0.0.1/sbin/restart-oms.sh

The following information is displayed:

```
start HA successfully.
```

Step 8 Log in to the standby management node and switch to user **omm**. Repeat step **Step 6** to step **Step 7**.

Run the **sh ${BIGDATA_HOME}/om-0.0.1/sbin/status-oms.sh** command to check whether **HAAllResOK** of the management node is **Normal**. Access MRS Manager again. If MRS Manager can be accessed, the operation is successful.

----End

10.11.12 Updating Cluster Keys

Scenario

When a cluster is installed, an encryption key is generated automatically to store the security information in the cluster (such as all database user passwords and key file access passwords) in encryption mode. After the cluster is successfully installed, you are advised to periodically update the encryption key based on the following procedure.

Impact on the System

- After a cluster key is updated, a new key is generated randomly in the cluster. This key is used to encrypt and decrypt the newly stored data. The old key is not deleted, and it is used to decrypt data encrypted using the old key. After security information is modified, for example, a database user password is changed, the new password is encrypted using the new key.
- When the key is updated, the cluster is stopped and cannot be accessed.

Prerequisites

The upper-layer applications depending on the cluster are stopped.

Procedure

Step 1 Log in to MRS Manager and choose **Services** > **More** > **Stop Cluster**.

In the displayed dialog box, select **I have read the information and understand the impact.** Click **OK**. Wait until the system displays a message indicating that the operation is successful. Click **Finish**. The cluster is stopped successfully.

Step 2 Log in to the active management node.

Step 3 Run the following commands to switch the user:

sudo su - omm

Step 4 Run the following command to disable logout upon timeout:

TMOUT=0

Step 5 Run the following command to switch the directory:

cd ${BIGDATA_HOME}/om-0.0.1/tools

Step 6 Run the following command to update the cluster key:

sh updateRootKey.sh

Enter **y** as prompted.

```
The root key update is a critical operation.
Do you want to continue?(y/n):
```

The key is updated successfully if the following information is displayed:

```
...
Step 4-1: The key save path is obtained successfully.
...
Step 4-4: The root key is sent successfully.
```

Step 7 On MRS Manager, choose **Services > More > Start Cluster**.

In the displayed dialog box, click **OK**. After **Operation successful** is displayed, click **Finish**. The cluster is started.

----End

10.12 Permissions Management

10.12.1 Creating a Role

Scenario

This section describes how to create a role on MRS Manager and authorize and manage Manager and components.

Up to 1,000 roles can be created on MRS Manager.

Prerequisites

You have learned service requirements.

Procedure

Step 1 On MRS Manager, choose **System** > **Manage Role**.

Step 2 Click **Create Role** and fill in **Role Name** and **Description**.

Role Name is mandatory and contains 3 to 30 digits, letters, and underscores (_). **Description** is optional.

Step 3 In **Permission**, set role permission.

1. Click **Service Name** and select a name in **View Name**.
2. Select one or more permissions.

NOTE

- The **Permission** parameter is optional.
- If you select **View Name** to set component permissions, you can enter a resource name in the **Search** box in the upper right corner and click 🔍. The search result is displayed.
- The search scope covers only directories with current permissions. You cannot search subdirectories. Search by keywords supports fuzzy match and is case-insensitive. Results of the next page can be searched.

Table 10-36 Manager permission description

Resource Supporting Permission Management	Permission Setting
Alarm	Authorizes the Manager alarm function. You can select **View** to view alarms and **Management** to manage alarms.
Audit	Authorizes the Manager audit log function. You can select **View** to view audit logs and **Management** to manage audit logs.
Dashboard	Authorizes the Manager overview function. You can select **View** to view the cluster overview.
Hosts	Authorizes the node management function. You can select **View** to view node information and **Management** to manage nodes.
Services	Authorizes the service management function. You can select **View** to view service information and **Management** to manage services.
System_cluster_management	Authorizes the MRS cluster management function. You can select **Management** to use the MRS patch management function.
System_configuration	Authorizes the MRS cluster configuration function. You can select **Management** to configure MRS clusters on Manager.
System_task	Authorizes the MRS cluster task function. You can select **Management** to manage periodic tasks of MRS clusters on Manager.
Tenant	Authorizes the Manager multi-tenant management function. You can select **Management** to manage multi-tenants.

Table 10-37 HBase permission description

Resource Supporting Permission Management	Permission Setting
SUPER_USER_GROUP	Grants you HBase administrator rights.
Global	HBase resource type, indicating the whole HBase.
Namespace	HBase resource type, indicating namespace, which is used to store HBase tables. It has the following permissions: ● **Admin** permission to manage the namespace ● **Create**: permission to create HBase tables in the namespace ● **Read**: permission to access the namespace ● **Write**: permission to write data to the namespace ● **Execute**: permission to execute the coprocessor (Endpoint)
Table	HBase resource type, indicating a data table, which is used to store data. It has the following permissions: ● **Admin**: permission to manage a data table ● **Create**: permission to create column families and columns in a data table ● **Read**: permission to read a data table ● **Write**: permission to write data to a data table ● **Execute**: permission to execute the coprocessor (Endpoint)
ColumnFamily	HBase resource type, indicating a column family, which is used to store data. It has the following permissions: ● **Create**: permission to create columns in a column family ● **Read**: permission to read a column family ● **Write**: permission to write data to a column family
Qualifier	HBase resource type, indicating a column, which is used to store data. It has the following permissions: ● **Read**: permission to read a column ● **Write**: permission to write data to a column

By default, permissions of an HBase resource type of each level are shared by resource types of sub-levels. However, the **Recursive** option is not selected by default. For example, if **Read** and **Write** permissions are added to the **default** namespace, they are automatically added to the tables, column families, and columns in the namespace. If a child resource is set after the parent resource, the permission of the child resource is the union of the permissions of the parent resource and the current child resource.

Table 10-38 HDFS permission description

Resource Supporting Permission Management	Permission Setting
Folder	HDFS resource type, indicating an HDFS directory, which is used to store files or subdirectories. It has the following permissions: ● **Read**: permission to access the HDFS directory ● **Write**: permission to write data to the HDFS directory ● **Execute**: permission to perform an operation. It must be selected when you add access or write permission.
Files	HDFS resource type, indicating a file in HDFS. It has the following permissions: ● **Read**: permission to access the file ● **Write**: permission to write data to the file ● **Execute**: permission to perform an operation. It must be selected when you add access or write permission.

Permissions of an HDFS directory of each level are not shared by directory types of sub-levels by default. For example, if **Read** and **Execute** permissions are added to the **tmp** directory, you must select **Recursive** at the same time to add permissions to subdirectories.

Table 10-39 Hive permission description

Resource Supporting Permission Management	Permission Setting
Hive Admin Privilege	Grants you Hive administrator rights.

Resource Supporting Permission Management	Permission Setting
Database	Hive resource type, indicating a Hive database, which is used to store Hive tables. It has the following permissions: ● **Select**: permission to query the Hive database ● **Delete**: permission to perform the deletion operation in the Hive database ● **Insert**: permission to perform the insertion operation in the Hive database ● **Create**: permission to perform the creation operation in the Hive database
Table	Hive resource type, indicating a Hive table, which is used to store data. It has the following permissions: ● **Select**: permission to query the Hive table ● **Delete**: permission to perform the deletion operation in the Hive table ● **Update**: grants users the **Update** permission of the Hive table ● **Insert**: permission to perform the insertion operation in the Hive table ● **Grant of Select**: permission to grant the **Select** permission to other users using Hive statements ● **Grant of Delete**: permission to grant the **Delete** permission to other users using Hive statements ● **Grant of Update**: permission to grant the **Update** permission to other users using Hive statements ● **Grant of Insert**: permission to grant the **Insert** permission to other users using Hive statements

By default, permissions of a Hive resource type of each level are shared by resource types of sub-levels. However, the **Recursive** option is not selected by default. For example, if **Select** and **Insert** permissions are added to the **default** database, they are automatically added to the tables and columns in the database. If a child resource is set after the parent resource, the permission of the child resource is the union of the permissions of the parent resource and the current child resource.

Table 10-40 Yarn permission description

Resource Supporting Permission Management	Permission Setting
Cluster Admin Operations	Grants you Yarn administrator rights.
root	Root queue of Yarn. It has the following permissions: ● **Submit**: permission to submit jobs in the queue ● **Admin**: permission to manage permissions of the current queue
Parent Queue	Yarn resource type, indicating a parent queue containing sub-queues. A root queue is a type of a parent queue. It has the following permissions: ● **Submit**: permission to submit jobs in the queue ● **Admin**: permission to manage permissions of the current queue
Leaf Queue	Yarn resource type, indicating a leaf queue. It has the following permissions: ● **Submit**: permission to submit jobs in the queue ● **Admin**: permission to manage permissions of the current queue

By default, permissions of a Yarn resource type of each level are shared by resource types of sub-levels. However, the **Recursive** option is not selected by default. For example, if the **Submit** permission is added to the **root** queue, it is automatically added to the sub-queue. Permissions inherited by sub-queues will not be displayed as selected in the **Permission** table. If a child resource is set after the parent resource, the permission of the child resource is the union of the permissions of the parent resource and the current child resource.

Table 10-41 Hue permission description

Resource Supporting Permission Management	Permission Setting
Storage Policy Admin	Grants you storage policy administrator rights.

Step 4 Click **OK**. Return to **Manage Role**.

----End

Related Tasks

Modifying a role

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage Role**.

Step 3 In the row of the role to be modified, click **Modify** to modify role information.

NOTE

If you change permissions assigned by the role, it takes 3 minutes to make new configurations take effect.

Step 4 Click **OK**. The modification is complete.

----End

Deleting a role

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage Role**.

Step 3 In the row of the role to be deleted, click **Delete**.

Step 4 Click **OK**. The role is deleted.

----End

10.12.2 Creating a User Group

Scenario

This section describes how to create user groups and specify their operation permissions on MRS Manager. Management of single or multiple users can be unified in the user groups. After being added to a user group, users can obtain operation permissions owned by the user group.

Up to 100 user groups can be created on MRS Manager.

Prerequisites

You have learned service requirements and created roles required by service scenarios.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage User Group**.

Step 3 Above the user group list, click **Create User Group**.

Step 4 Input **Group Name** and **Description**.

Group Name is mandatory and contains 3 to 20 digits, letters, and underscores (_). **Description** is optional.

Step 5 In **Role**, click **Select and Add Role** to select and add specified roles.

If you do not add the roles, the user group you are creating now does not have the permission to use MRS clusters.

Step 6 Click **OK**. The user group is created.

----End

Related Tasks

Modifying a user group

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage User Group**.

Step 3 In the row of the user group to be modified, click **Modify**.

NOTE

If you change role permissions assigned to the user group, it takes 3 minutes to make new configurations take effect.

Step 4 Click **OK**. The modification is complete.

----End

Deleting a user group

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage User Group**.

Step 3 In the row of the user group to be deleted, click **Delete**.

Step 4 Click **OK**. The user group is deleted.

----End

10.12.3 Creating a User

Scenario

This section describes how to create users on MRS Manager based on site requirements and specify their operation permissions to meet service requirements.

Up to 1,000 users can be created on MRS Manager.

If a new password policy needs to be used for a new user's password, follow instructions in **Modifying a Password Policy** to modify the password policy and then perform the following operations to create a user.

Prerequisites

You have learned service requirements and created roles and role groups required by service scenarios.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage User**.

Step 3 Above the user list, click **Create User**.

Step 4 Configure parameters as prompted and enter a username in **User Name**.

NOTE

- If a username exists, you cannot create another username that only differs from the existing username in case. For example, if **User1** has been created, you cannot create **user1**.
- When you use the user you created, enter the correct username, which is case-sensitive.
- **User Name** is mandatory and contains 3 to 20 digits, letters, and underscores (_).
- **root**, **omm**, and **ommdba** are reserved system user. Select another username.

Step 5 Set **User Type** to either **Human-Machine** or **Machine-Machine**.

- **Human-Machine** users: used for O&M on MRS Manager and operations on component clients. If you select this user type, you need to enter a password and confirm the password in **Password** and **Confirm Password** accordingly.
- **Machine-Machine** users: used for MRS application development. If you select this user type, you do not need to enter a password, because the password is randomly generated.

Step 6 In **User Group**, click **Select and Join User Group** to select user groups and add users to them.

NOTE

- If roles have been added to user groups, the users can be granted with permissions of the roles.
- If you want to grant new users with Hive permissions, add the users to the Hive group.
- If a user needs to manage tenant resources, the user group must be assigned the **Manager_tenant** role and the role corresponding to the tenant.

Step 7 In **Primary Group**, select a group as the primary group for users to create directories and files. The drop-down list contains all groups selected in **User Group**.

Step 8 In **Assign Rights by Role**, click **Select and Add Role** to add roles for users based on service requirements.

NOTE

- When you create a user, if permissions of a user group that is granted to the user cannot meet service requirements, you can assign other created roles to the user. It takes 3 minutes to make role permissions granted to the new user take effect.
- Adding a role when you create a user can specify the user rights.
- A new user can access WebUIs of HDFS, HBase, Yarn, Spark, and Hue even when roles are not assigned to the user.

Step 9 In **Description**, provide description based on onsite service requirements.

Description is optional.

Step 10 Click **OK**. The user is created.

If a new user is used in the MRS cluster for the first time, for example, used for logging in to MRS Manager or using the cluster client, the password must be changed. For details, see section **Changing the Password of an Operation User**.

----End

10.12.4 Modifying User Information

Scenario

This section describes how to modify user information on MRS Manager, including information about the user group, primary group, role, and description.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage User**.

Step 3 In the row of the user to be modified, click **Modify**.

NOTE

If you change user groups for or assign role permissions to the user, it takes 3 minutes to make new configurations take effect.

Step 4 Click **OK**. The modification is complete.

----End

10.12.5 Locking a User

This section describes how to lock users in MRS clusters. A locked user cannot log in to MRS Manager or perform security authentication in the cluster.

A locked user can be unlocked by manually or until the lock duration expires. You can lock a user by using either of the following methods:

- Automatic lock: Set **Number of Password Retries** in **Configure Password Policy**. If user login attempts exceed the parameter value, the user is automatically locked. For details, see **Modifying a Password Policy** .
- Manual lock: manually locks a user.

The following describes how to manually lock a user. **Machine-Machine** users cannot be locked.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage User**.

Step 3 In the row of the user to be locked, click **Lock User**.

Step 4 In the window that is displayed, click **Yes** to lock the user.

----End

10.12.6 Unlocking a User

If a user is locked because the number of login attempts exceeds the value of **Number of Password Retries**, or the user is manually locked, the administrator can unlock the user on MRS Manager.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage User**.

Step 3 In the row of the user to be unlocked, click **Unlock User**.

Step 4 In the window that is displayed, click **Yes** to unlock the user.

----End

10.12.7 Deleting a User

Scenario

If an MRS cluster user is not required, you can delete the user on MRS Manager.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage User**.

Step 3 In the row of the user to be deleted, choose **More** > **Delete**.

Step 4 Click **OK**.

----End

10.12.8 Changing the Password of an Operation User

Scenario

Passwords of **Human-Machine** system users must be regularly changed to ensure MRS cluster security. This section describes how to change your passwords on MRS Manager.

If a new password policy needs to be used for the password modified by the user, follow instructions in **Modifying a Password Policy** to modify the password policy and then perform the following operations to modify the password.

Impact on the System

If you have downloaded a user authentication file, download it again and obtain the keytab file after changing the password of the MRS cluster user.

Prerequisites

- You have obtained the current password policies.
- You have obtained the MRS Manager access address.

Procedure

Step 1 On MRS Manager, move the mouse cursor to in the upper right corner.

On the menu that is displayed, select **Change Password**.

Step 2 Fill in the **Old Password**, **New Password**, and **Confirm Password**. Click **OK**.

For the cluster, the default password complexity requirements are as follows:

- The password must contain 8 to 32 characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, spaces, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The password cannot be the username or the reverse username.

----End

10.12.9 Initializing the Password of a System User

Scenario

This section describes how to initialize a password on MRS Manager if a user forgets the password or the password of a public account needs to be changed regularly. After password initialization, the user must change the password upon the first login.

Impact on the System

If you have downloaded a user authentication file, download it again and obtain the keytab file after initializing the password of the MRS cluster user.

Initializing the Password of a Human-Machine User

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage User**.

Step 3 Locate the row that contains the user whose password is to be initialized, choose **More** > **Initialize password**, and change the password as prompted.

In the window that is displayed, enter the password and click **OK**. Then in **Initialize password**, click **OK**.

For the cluster, the default password complexity requirements are as follows:

- The password must contain 8 to 32 characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, spaces, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).

- The password cannot be the username or the reverse username.

----End

Initializing the Password of a Machine-Machine User

Step 1 Prepare a client based on service conditions and log in to the node where the client is installed.

Step 2 Run the following command to switch the user:

sudo su - omm

Step 3 Run the following command to switch to the client directory, for example, **/opt/client**:

cd /opt/client

Step 4 Run the following command to configure environment variables:

source bigdata_env

Step 5 Run the following command to log in to the console as user **kadmin/admin**:

kadmin -p kadmin/admin

NOTE

The default password of user **kadmin/admin** is **KAdmin@123**, which will expire upon your first login. Change the password as prompted and keep the new password secure.

Step 6 Run the following command to reset the password of a component running user. This operation takes effect for all servers.

cpw *Component running user name*

For example, **cpw oms/manager**.

For the cluster, the default password complexity requirements are as follows:

- The password must contain 8 to 32 characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, spaces, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The password cannot be the username or the reverse username.

----End

10.12.10 Downloading a User Authentication File

Scenario

When a user develops big data applications and runs them in an MRS cluster that supports Kerberos authentication, the user needs to prepare a user authentication file for accessing the MRS cluster. The keytab file in the authentication file can be used for user authentication.

This section describes how to download a user authentication file and export the keytab file on MRS Manager.

NOTE

- Before downloading a **Human-machine** user authentication file, change the password for the user on MRS Manager to make the initial password set by the administrator invalid. Otherwise, the exported keytab file cannot be used. For details, see **Changing the Password of an Operation User**.
- After a user password is changed, the exported keytab file becomes invalid, and you need to export a keytab file again.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage User**.

Step 3 In the row of the user for whom you want to export the keytab file, choose **More** > **Download authentication credential** to download the authentication file. After the file is automatically generated, save it to a specified path and keep it properly.

Step 4 Open the authentication file with a decompression program.

- **user.keytab** indicates a user keytab file used for user authentication.
- **krb5.conf** indicates the configuration file of the authentication server. The application connects to the authentication server according to the configuration file information when authenticating users.

----End

10.12.11 Modifying a Password Policy

Scenario

This section describes how to set password and user login security rules as well as user lock rules. Password policies set on MRS Manager take effect for **Human-machine** users only, because the passwords of **Machine-machine** users are randomly generated.

If a new password policy needs to be used for a new user's password or the password modified by the user, perform the following operations to modify the password policy first, and then create a user or change the password by following instructions in **Creating a User** or **Changing the Password of an Operation User**.

NOTICE

Modify password policies based on service security requirements, because they involve user management security. Otherwise, security risks may be caused.

Procedure

Step 1 On MRS Manager, click **System**.

Step 2 Click **Configure Password Policy**.

Step 3 Modify password policies as prompted. For parameter details, see the following table:

Table 10-42 Password policy parameter description

Parameter	Description
Minimum Password Length	Indicates the minimum number of characters a password contains. The value ranges from 8 to 32. The default value is **8**.
Number of Character Types	Indicates the minimum number of character types a password contains. The character types are uppercase letters, lowercase letters, digits, spaces, and special characters (~`!?,.:;-_'(){}[]/<>@#$%^&*+|\=). The value can be **3** or **4**. The default value **3** indicates that the password must contain at least three types of the following characters: uppercase letters, lowercase letters, digits, special characters, and spaces.
Password Validity Period (days)	Indicates the validity period (days) of a password. The value ranges from 0 to 90. 0 means that the password is permanently valid. The default value is **90**.
Password Expiration Notification Days	Indicates the number of days in advance users are notified that their passwords are about to expire. After the value is set, if the difference between the cluster time and the password expiration time is smaller than this value, the user receives password expiration notifications. When a user logs in to MRS Manager, a message is displayed, indicating that the password is about to expire and asking the user whether to change the password. The value ranges from **0** to *X* (*X* must be set to the half of the password validity period and rounded down). Value **0** indicates that no notification is sent. The default value is **5**.
Interval of Resetting Authentication Failure Count (min)	Indicates the interval of retaining incorrect password attempts, in minutes. The value ranges from 0 to 1440. 0 indicates that incorrect password attempts are permanently retained and 1440 indicates that incorrect password attempts are retained for one day. The default value is **5**.
Number of Password Retries	Indicates the number of consecutive wrong passwords allowed before the system locks the user. The value ranges from 3 to 30. The default value is **5**.

Parameter	Description
Account Lock Duration (min)	Indicates the time period for which a user is locked when the user lockout conditions are met. The value ranges from 5 to 120. The default value is **5**.

----End

10.13 MRS Multi-User Permission Management

10.13.1 Users and Permissions of MRS Clusters

Overview

- **MRS Cluster Users**

  Indicate the security accounts of Manager, including usernames and passwords. These accounts are used to access resources in MRS clusters. Each MRS cluster in which Kerberos authentication is enabled can have multiple users.

- **MRS Cluster Roles**

  Before using resources in an MRS cluster, users must obtain the access permission which is defined by MRS cluster objects. A cluster role is a set of one or more permissions. For example, the permission to access a directory in HDFS needs to be configured in the specified directory and saved in a role.

Manager provides the user permission management function for MRS clusters, facilitating permission and user management.

- Permission management: adopts the role-based access control (RBAC) mode. In this mode, permissions are granted by role to form a permission set. After one or more roles are allocated to a user, the user can obtain the permissions of the roles.
- User management: uses MRS Manager to uniformly manage users, adopts the Kerberos protocol for user identity verification, and employs Lightweight Directory Access Protocol (LDAP) to store user information.

Permission Management

Permissions provided by MRS clusters include the O&M permissions of Manager and components (such as HDFS, HBase, Hive, and Yarn). In actual application, permissions must be assigned to each user based on service scenarios. To facilitate permission management, Manager introduces the role function to allow administrators to select and assign specified permissions. Permissions are centrally viewed and managed in permission sets, enhancing user experience.

A role is a logical entity that contains one or more permissions. Permissions are assigned to roles, and users can be granted the permissions by obtaining the roles.

A role can have multiple permissions, and a user can be bound to multiple roles.

- Role 1: is assigned operation permissions A and B. After role 1 is allocated to users a and b, users a and b can obtain operation permissions A and B.
- Role 2: is assigned operation permission C. After role 2 is allocated to users c and d, users c and d can obtain operation permission C.
- Role 3: is assigned operation permissions D and F. After role 3 is allocated to user a, user a can obtain operation permissions D and F.

For example, if an MRS user is bound to the administrator role, the user becomes an administrator of the MRS cluster.

Table 10-43 lists the roles that are created by default on Manager.

Table 10-43 Default roles and description

Default Role	Description
default	Tenant role
Manager_administrator	Manager administrator: This role has the permission to manage MRS Manager.
Manager_auditor	Manager auditor: This role has the permission to view and manage auditing information.
Manager_operator	Manager operator: This role has all permissions except tenant, configuration, and cluster management permissions.
Manager_viewer	Manager viewer: This role has the permission to view the information about systems, services, hosts, alarms, and auditing logs.
System_administrator	System administrator: This role has the permissions of Manager administrators and all service administrators.
Manager_tenant	Manager tenant viewer: This role has the permission to view information on the **Tenant** page on MRS Manager.

When creating a role on Manager, you can perform rights management for Manager and components, as shown in **Table 10-44**.

Table 10-44 Manager and component permission management

Permission	Description
Manager	Manager access and login permission.
HBase	HBase administrator permission and permission for accessing HBase tables and column families.
HDFS	HDFS directory and file permission.

Permission	Description
Hive	- Hive Admin Privilege Hive administrator permission. - Hive Read Write Privileges Hive data table management permission to set and manage the data of created tables.
Hue	Storage policy administrator permissions.
Yarn	- Cluster Admin Operations Yarn administrator permission. - Scheduler Queue Queue resource management permission.

User Management

MRS clusters that support Kerberos authentication use the Kerberos protocol and LDAP for user management.

- Kerberos verifies the identity of the user when a user logs in to Manager or uses a component client. Identity verification is not required for clusters with Kerberos authentication disabled.
- LDAP is used to store user information, including user records, user group information, and permission information.

MRS clusters can automatically update Kerberos and LDAP user data when users are created or modified on Manager. They can also automatically perform user identity verification and authentication and obtain user information when a user logs in to Manager or uses a component client. This ensures the security of user management and simplifies the user management tasks. Manager also provides the user group function for managing one or multiple users by type:

- A user group is a set of users, which can be used to manage users by type. Users in the system can exist independently or in a user group.
- After a user is added to a user group to which roles are allocated, the role permission of the user group is assigned to the user.

Table 10-45 lists the user groups that are created by default on MRS Manager in MRS 3.x or earlier.

For details about the default user groups displayed on FusionInsight Manager of MRS 3.*x* or later, see **User group**.

Table 10-45 Default user groups and description

User Group	Description
hadoop	Users added to this user group have the permission to submit tasks to all Yarn queues.

User Group	Description
hbase	Common user group. Users added to this user group will not have any additional permission.
hive	Users added to this user group can use Hive.
spark	Common user group. Users added to this user group will not have any additional permission.
supergroup	Users added to this user group can have the administrator permission of HBase, HDFS, and Yarn and can use Hive.
flume	Common user group. Users added to this user group will not have any additional permission.
kafka	Kafka common user group. Users added to this group need to be granted with read and write permission by users in the **kafkaadmin** group before accessing the desired topics.
kafkasuperuser	Users added to this group have permissions to read data from and write data to all topics.
kafkaadmin	Kafka administrator group. Users added to this group have the permissions to create, delete, authorize, as well as read from and write data to all topics.
storm	Storm common user group. Users added to this group have the permissions to submit topologies and manage their own topologies.
stormadmin	Storm administrator user group. Users added to this group have the permissions to submit topologies and manage their own topologies.

User **admin** is created by default for MRS clusters with Kerberos authentication enabled and is used for administrators to maintain the clusters.

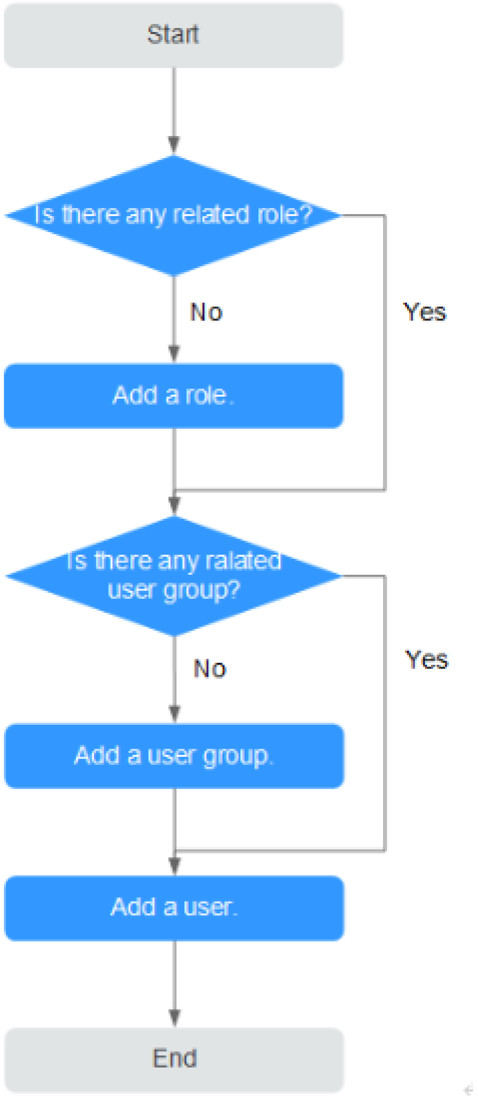
Process Overview

In practice, MRS cluster users must understand the service scenarios of big data and plan user permissions. Then, create roles and assign permissions to the roles on MRS Manager to meet service requirements. Manager provides the user group function for administrators to create user groups for managing users of one or multiple service scenarios of the same type.

NOTE

If a role has the permission of HDFS, HBase, Hive, or Yarn respectively, the role can only use the corresponding functions of the component. To use Manager, the corresponding Manager permission must be added to the role.

Figure 10-1 Process of creating a user



10.13.2 Default Users of Clusters with Kerberos Authentication Enabled

User Classification

The MRS cluster provides the following three types of users. Users are advised to periodically change the passwords. It is not recommended to use the default passwords.

User Type	Description
System user	● User created on Manager for MRS cluster O&M and service scenarios. There are two types of users: – **Human-machine** user: used for Manager O&M scenarios and component client operation scenarios. – **Machine-machine** user: used for MRS cluster application development scenarios. ● User who runs OMS processes.
Internal system user	Internal user who performs process communications, saves user group information, and associates user permissions.
Database user	● User who manages OMS database and accesses data. ● User who runs the database of service components (Hive, Hue, Loader, and DBService)

System User

NOTE

- User **ldap** of the OS is required in the MRS cluster. Do not delete this account. Otherwise, the cluster may not work properly. Password management policies are maintained by the operation users.
- Reset the passwords when you change the passwords of user **ommdba** and user **omm** for the first time. Change the passwords periodically after retrieving them.

Type	Username	Initial Password	Description
System administrator of the MRS cluster	admin	Specified by the user during the cluster creation.	Manager administrator with the following permissions: ● Common HDFS and ZooKeeper user permissions. ● Permissions to submit and query MapReduce and Yarn tasks, manage Yarn queues, and access the Yarn web UI. ● Permissions to submit, query, activate, deactivate, reassign, delete topologies, and operate all topologies of the Storm service. ● Permissions to create, delete, authorize, reassign, consume, write, and query topics of the Kafka service.
MRS cluster node OS user	omm	Randomly generated by the system.	Internal running user of the MRS cluster system. This user is an OS user generated on all nodes and does not require a unified password.
MRS cluster node OS user	root	Set by the user.	User for logging in to the node in the MRS cluster. This user is an OS user generated on all nodes.

Internal System Users

NOTE

Do not delete the following internal system users. Otherwise, the cluster or components may not work properly.

Type	Default User	Initial Password	Description
Component running user	hdfs	Hdfs@123	This user is the HDFS system administrator and has the following permissions: 1. File system operation permissions: ● Views, modifies, and creates files. ● Views and creates directories. ● Views and modifies the groups where files belong. ● Views and sets disk quotas for users. 2. HDFS management operation permissions: ● Views the web UI status. ● Views and sets the active and standby HDFS status. ● Enters and exits the HDFS in security mode. ● Checks the HDFS file system.

Type	Default User	Initial Password	Description
	hbase	Hbase@123	This user is the HBase system administrator and has the following permissions: ● Cluster management permission: **Enable** and **Disable** operations on tables to trigger MajorCompact and ACL operations. ● Grants and revokes permissions, and shuts down the cluster. ● Table management permission: Creates, modifies, and deletes tables. ● Data management permission: Reads and writes data in tables, column families, and columns. ● Accesses the HBase web UI.
	mapred	Mapred@123	This user is the MapReduce system administrator and has the following permissions: ● Submits, stops, and views the MapReduce tasks. ● Modifies the Yarn configuration parameters. ● Accesses the Yarn and MapReduce web UI.
	spark	Spark@123	This user is the Spark system administrator and has the following permissions: ● Accesses the Spark web UI. ● Submits Spark tasks.

User Group Information

Default User Group	Description
hadoop	Users added to this user group have the permission to submit tasks to all Yarn queues.
hbase	Common user group. Users added to this user group will not have any additional permission.
hive	Users added to this user group can use Hive.
spark	Common user group. Users added to this user group will not have any additional permission.
supergroup	Users added to this user group can have the administrator permission of HBase, HDFS, and Yarn and can use Hive.
check_sec_ldap	Used to test whether the active LDAP works properly. This user group is generated randomly in a test and automatically deleted after the test is complete. This is an internal system user group used only between components.
Manager_tenant	Tenant system user group, which is an internal system user group used only between components.
System_administrator	MRS cluster system administrator group, which is an internal system user group used only between components.
Manager_viewer	MRS Manager system viewer group, which is an internal system user group used only between components.
Manager_operator	MRS Manager system operator group, which is an internal system user group used only between components.
Manager_auditor	MRS Manager system auditor group, which is an internal system user group used only between components.
Manager_administrator	MRS Manager system administrator group, which is an internal system user group used only between components.
compcommon	Internal system group for accessing public resources in a cluster. All system users and system running users are added to this user group by default.
default_1000	User group created for tenants, which is an internal system user group used only between components.

Default User Group	Description
kafka	Kafka common user group. Users added to this group need to be granted with read and write permission by users in the **kafkaadmin** group before accessing the desired topics.
kafkasuperuser	Users added to this group have permissions to read data from and write data to all topics.
kafkaadmin	Kafka administrator group. Users added to this group have the permissions to create, delete, authorize, as well as read from and write data to all topics.
storm	Storm common user group. Users added to this group have the permissions to submit topologies and manage their own topologies.
stormadmin	Storm administrator user group. Users added to this group have the permissions to submit topologies and manage their own topologies.
opentsdb	Common user group. Users added to this user group will not have any additional permission.
presto	Common user group. Users added to this user group will not have any additional permission.
flume	Common user group. Users added to this user group will not have any additional permission.
launcher-job	MRS internal group, which is used to submit jobs using V2 APIs.

OS User Group	Description
wheel	Primary group of MRS internal running user **omm**.
ficommon	MRS cluster common group that corresponds to **compcommon** for accessing public resource files stored in the OS of the cluster.

Database User

MRS cluster system database users include OMS database users and DBService database users.

NOTE

Do not delete database users. Otherwise, the cluster or components may not work properly.

Type	Default User	Initial Password	Description
OMS database	ommdba	dbChangeMe@123456	OMS database administrator who performs maintenance operations, such as creating, starting, and stopping applications.
	omm	ChangeMe@123456	User for accessing OMS database data.
DBService database	omm	dbserverAdmin@123	Administrator of the GaussDB database in the DBService component.
	hive	HiveUser@	User for Hive to connect to the DBService database.
	hue	HueUser@123	User for Hue to connect to the DBService database.
	sqoop	SqoopUser@	User for Loader to connect to the DBService database.
	ranger	RangerUser@	User for Ranger to connect to the DBService database.

10.13.3 Creating a Role

Scenario

This section describes how to create a role on Manager and authorize and manage Manager and components.

Up to 1000 roles can be created on Manager.

NOTE

The operations described in this section apply only to clusters of versions earlier than MRS 3.x.

For clusters of **MRS 3.*x*** or later, see **Managing Roles**.

Prerequisites

- You have learned service requirements.
- You have obtained a cluster with Kerberos authentication enabled or a common cluster with the EIP function enabled.

Procedure

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 On MRS Manager, choose **System** > **Manage Role**.

Step 3 Click **Create Role** and fill in **Role Name** and **Description**.

Role Name is mandatory and contains 3 to 30 characters. Only digits, letters, and underscores (_) are allowed. **Description** is optional.

Step 4 In **Permission**, set role permission.

1. Click **Service Name** and select a name in **View Name**.
2. Select one or more permissions.

NOTE

- The **Permission** parameter is optional.
- If you select **View Name** to set component permissions, you can enter a resource name in the **Search** box in the upper right corner and click 🔍. The search result is displayed.
- The search scope covers only directories with current permissions. You cannot search subdirectories. Search by keywords supports fuzzy match and is case-insensitive. Results of the next page can be searched.

Table 10-46 Manager permission description

Resource Supporting Permission Management	Permission Setting
Alarm	Authorizes the Manager alarm function. You can select **View** to view alarms and **Management** to manage alarms.
Audit	Authorizes the Manager audit log function. You can select **View** to view audit logs and **Management** to manage audit logs.
Dashboard	Authorizes the Manager overview function. You can select **View** to view the cluster overview.
Hosts	Authorizes the node management function. You can select **View** to view node information and **Management** to manage nodes.
Services	Authorizes the service management function. You can select **View** to view service information and **Management** to manage services.
System_cluster_management	Authorizes the MRS cluster management function. You can select **Management** to use the MRS patch management function.
System_configuration	Authorizes the MRS cluster configuration function. You can select **Management** to configure MRS clusters on Manager.
System_task	Authorizes the MRS cluster task function. You can select **Management** to manage periodic tasks of MRS clusters on Manager.

Resource Supporting Permission Management	Permission Setting
Tenant	Authorizes the Manager multi-tenant management function. You can select **Management** to manage multi-tenants.

Table 10-47 HBase permission description

Resource Supporting Permission Management	Permission Setting
SUPER_USER_GROUP	Grants you HBase administrator permissions.
Global	HBase resource type, indicating the whole HBase.
Namespace	HBase resource type, indicating namespace, which is used to store HBase tables. It has the following permissions: ● **Admin** permission to manage the namespace ● **Create**: permission to create HBase tables in the namespace ● **Read**: permission to access the namespace ● **Write**: permission to write data to the namespace ● **Execute**: permission to execute the coprocessor (Endpoint)
Table	HBase resource type, indicating a data table, which is used to store data. It has the following permissions: ● **Admin**: permission to manage a data table ● **Create**: permission to create column families and columns in a data table ● **Read**: permission to read a data table ● **Write**: permission to write data to a data table ● **Execute**: permission to execute the coprocessor (Endpoint)
ColumnFamily	HBase resource type, indicating a column family, which is used to store data. It has the following permissions: ● **Create**: permission to create columns in a column family ● **Read**: permission to read a column family ● **Write**: permission to write data to a column family

Resource Supporting Permission Management	Permission Setting
Qualifier	HBase resource type, indicating a column, which is used to store data. It has the following permissions: ● **Read**: permission to read a column ● **Write**: permission to write data to a column

By default, permissions of an HBase resource type of each level are shared by resource types of sub-levels. However, the **Recursive** option is not selected by default. For example, if **Read** and **Write** permissions are added to the **default** namespace, they are automatically added to the tables, column families, and columns in the namespace. If a child resource is set after the parent resource, the permission of the child resource is the union of the permissions of the parent resource and the current child resource.

Table 10-48 HDFS permission description

Resource Supporting Permission Management	Permission Setting
Folder	HDFS resource type, indicating an HDFS directory, which is used to store files or subdirectories. It has the following permissions: ● **Read**: permission to access the HDFS directory ● **Write**: permission to write data to the HDFS directory ● **Execute**: permission to perform an operation. It must be selected when you add access or write permission.
Files	HDFS resource type, indicating a file in HDFS. It has the following permissions: ● **Read**: permission to access the file ● **Write**: permission to write data to the file ● **Execute**: permission to perform an operation. It must be selected when you add access or write permission.

Permissions of an HDFS directory of each level are not shared by directory types of sub-levels by default. For example, if **Read** and **Execute** permissions are added to the **tmp** directory, you must select **Recursive** for permissions to be added to subdirectories.

Table 10-49 Hive permission description

Resource Supporting Permission Management	Permission Setting
Hive Admin Privilege	Grants you Hive administrator permissions.
Database	Hive resource type, indicating a Hive database, which is used to store Hive tables. It has the following permissions: ● **Select**: permission to query the Hive database ● **Delete**: permission to perform the deletion operation in the Hive database ● **Insert**: permission to perform the insertion operation in the Hive database ● **Create**: permission to perform the creation operation in the Hive database
Table	Hive resource type, indicating a Hive table, which is used to store data. It has the following permissions: ● **Select**: permission to query the Hive table ● **Delete**: permission to perform the deletion operation in the Hive table ● **Update**: permission to perform the update operation in the Hive table ● **Insert**: permission to perform the insertion operation in the Hive table ● **Grant of Select**: permission to grant the **Select** permission to other users using Hive statements ● **Grant of Delete**: permission to grant the **Delete** permission to other users using Hive statements ● **Grant of Update**: permission to grant the **Update** permission to other users using Hive statements ● **Grant of Insert**: permission to grant the **Insert** permission to other users using Hive statements

By default, permissions of a Hive resource type of each level are shared by resource types of sub-levels. However, the **Recursive** option is not selected by default. For example, if **Select** and **Insert** permissions are added to the **default** database, they are automatically added to the tables and columns in the database. If a child resource is set after the parent resource, the permission of the child resource is the union of the permissions of the parent resource and the current child resource.

Table 10-50 Yarn permission description

Resource Supporting Permission Management	Permission Setting
Cluster Admin Operations	Grants you Yarn administrator permissions.
root	Root queue of Yarn. It has the following permissions: ● **Submit**: permission to submit jobs in the queue ● **Admin**: permission to manage permissions of the current queue
Parent Queue	Yarn resource type, indicating a parent queue containing sub-queues. A root queue is a type of a parent queue. It has the following permissions: ● **Submit**: permission to submit jobs in the queue ● **Admin**: permission to manage permissions of the current queue
Leaf Queue	Yarn resource type, indicating a leaf queue. It has the following permissions: ● **Submit**: permission to submit jobs in the queue ● **Admin**: permission to manage permissions of the current queue

By default, permissions of a Yarn resource type of each level are shared by resource types of sub-levels. However, the **Recursive** option is not selected by default. For example, if the **Submit** permission is added to the **root** queue, it is automatically added to the sub-queue. Permissions inherited by sub-queues will not be displayed as selected in the **Permission** table. If a child resource is set after the parent resource, the permission of the child resource is the union of the permissions of the parent resource and the current child resource.

Table 10-51 Hue permission description

Resource Supporting Permission Management	Permission Setting
Storage Policy Admin	Grants you storage policy administrator permissions.

Step 5 Click **OK**. Return to **Manage Role**.

----End

Related Tasks

Modifying a role

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage Role**.

Step 3 In the row of the role to be modified, click **Modify** to modify role information.

NOTE

If you modify permissions assigned by the role, it takes 3 minutes to make new configurations take effect.

Step 4 Click **OK**. The modification is complete.

----End

Deleting a role

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage Role**.

Step 3 In the row of the role to be deleted, click **Delete**.

Step 4 Click **OK**. The role is deleted.

----End

10.13.4 Creating a User Group

Scenario

This section describes how to create user groups and specify their operation permissions on Manager. Management of single or multiple users can be unified in the user groups. After being added to a user group, users can obtain operation permissions owned by the user group.

Manager supports a maximum of 100 user groups.

NOTE

The operations described in this section apply only to clusters of versions earlier than MRS 3.x.

For clusters of **MRS 3.*x*** or later, see **Managing User Groups**.

Prerequisites

- You have learned service requirements and created roles required by service scenarios.
- You have obtained a cluster with Kerberos authentication enabled or a common cluster with the EIP function enabled.

Procedure

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 On MRS Manager, click **System**.

Step 3 In the **Permission** area, click **Manage User Group**.

Step 4 Above the user group list, click **Create User Group**.

Step 5 Input **Group Name** and **Description**.

Group Name is mandatory and contains 3 to 20 characters. Only digits, letters, and underscores (_) are allowed. **Description** is optional.

Step 6 In **Role**, click **Select and Add Role** to select and add specified roles.

If you do not add the roles, the user group you are creating now does not have the permission to use MRS clusters.

Step 7 Click **OK**.

----End

Related Tasks

Modifying a user group

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage User Group**.

Step 3 In the row of a user group to be modified, click **Modify**.

NOTE

If you change role permissions assigned to the user group, it takes 3 minutes to make new configurations take effect.

Step 4 Click **OK**. The modification is complete.

----End

Deleting a user group

Step 1 On MRS Manager, click **System**.

Step 2 In the **Permission** area, click **Manage User Group**.

Step 3 In the row of the user group to be deleted, click **Delete**.

Step 4 Click **OK**. The user group is deleted.

----End

10.13.5 Creating a User

Scenario

This section describes how to create users on Manager based on site requirements and specify their operation permissions to meet service requirements.

Up to 1000 users can be created on Manager.

If a new password policy needs to be used for a new user's password, follow instructions in **Modifying a Password Policy** to modify the password policy and then perform the following operations to create a user.

NOTE

The operations described in this section apply only to clusters of versions earlier than MRS 3.x.

For clusters of **MRS 3.*x*** or later, see **Creating a User**.

Prerequisites

- You have learned service requirements and created roles and role groups required by service scenarios.
- You have obtained a cluster with Kerberos authentication enabled or a common cluster with the EIP function enabled.

Procedure

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 On MRS Manager, click **System**.

Step 3 In the **Permission** area, click **Manage User**.

Step 4 Above the user list, click **Create User**.

Step 5 Configure parameters as prompted and enter a username in **Username**.

NOTE

- A username that differs only in alphabetic case from an existing username is not allowed. For example, if **User1** has been created, you cannot create **user1**.
- When you use the user you created, enter the exactly correct username, which is case-sensitive.
- **Username** is mandatory and contains 3 to 20 characters. Only digits, letters, and underscores (_) are allowed.
- **root**, **omm**, and **ommdba** are reserved system user. Select another username.

Step 6 Set **User Type** to either **Human-machine** or **Machine-machine**.

- **Human-machine** user: used for MRS Manager O&M scenarios and component client operation scenarios. If you select this user type, you need to enter a password and confirm the password in **Password** and **Confirm Password** accordingly.
- **Machine-machine** users: used for MRS application development scenarios. If you select this user type, you do not need to enter a password, because the password is randomly generated.

Step 7 In **User Group**, click **Select and Join User Group** to select user groups and add users to them.

NOTE

- If roles have been added to user groups, the users can be granted with permissions of the roles.
- If you want to grant new users with Hive permissions, add the users to the Hive group.
- If a user needs to manage tenant resources, the user group must be assigned the **Manager_tenant** role and the role corresponding to the tenant.
- Users created on Manager cannot be added to the user group synchronized using the IAM user synchronization function.

Step 8 In **Primary Group**, select a group as the primary group for users to create directories and files. The drop-down list contains all groups selected in **User Group**.

Step 9 In **Assign Rights by Role**, click **Select and Add Role** to add roles for users based on onsite service requirements.

NOTE

- When you create a user, if permissions of a user group that is granted to the user cannot meet service requirements, you can assign other created roles to the user. It takes 3 minutes to make role permissions granted to the new user take effect.
- Adding a role when you create a user can specify the user rights.
- A new user can access web UIs of HDFS, HBase, Yarn, Spark, and Hue even when roles are not assigned to the user.

Step 10 In **Description**, provide description based on onsite service requirements.

Description is optional.

Step 11 Click **OK**.

If a new user is used in the MRS cluster for the first time, for example, used for logging in to MRS Manager or using the cluster client, the password must be changed. For details, see **Changing the Password of an Operation User**.

----End

10.13.6 Modifying User Information

Scenario

This section describes how to modify user information on Manager, including information about the user group, primary group, role, and description.

This operation is supported only in clusters with Kerberos authentication enabled or common clusters with the EIP function enabled.

NOTE

The operations described in this section apply only to clusters of versions earlier than MRS 3.x.

For clusters of **MRS 3.*x*** or later, see **Modifying User Information**.

Procedure

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 On MRS Manager, click **System**.

Step 3 In the **Permission** area, click **Manage User**.

Step 4 In the row of a user to be modified, click **Modify**.

NOTE

If you change user groups for a user or assign role permissions to a user, it takes 3 minutes to make new configurations take effect.

Step 5 Click **OK**. The modification is complete.

----End

10.13.7 Locking a User

This section describes how to lock users in MRS clusters. A locked user cannot log in to Manager or perform security authentication in the cluster. This operation is supported only in clusters with Kerberos authentication enabled or common clusters with the EIP function enabled.

A locked user can be unlocked by manually or until the lock duration expires. You can lock a user by using either of the following methods:

- Automatic lock: Set **Number of Password Retries** in **Configure Password Policy**. If user login attempts exceed the parameter value, the user is automatically locked. For details, see **Modifying a Password Policy**.
- Manual lock: manually locks a user.

NOTE

The operations described in this section apply only to clusters of versions earlier than MRS 3.x.

For clusters of **MRS 3.*x*** or later, see **Locking a User**.

The following describes how to manually lock a user. **Machine-machine** users cannot be locked.

Procedure

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 On MRS Manager, click **System**.

Step 3 In the **Permission** area, click **Manage User**.

Step 4 In the row of a user you want to lock, click **Lock User**.

Step 5 In the window that is displayed, click **OK** to lock the user.

----End

10.13.8 Unlocking a User

If a user is locked because the number of login attempts exceeds the value of **Number of Password Retries**, or the user is manually locked, you can unlock the user on Manager. This operation is supported only in clusters with Kerberos authentication enabled or common clusters with the EIP function enabled.

NOTE

The operations described in this section apply only to clusters of versions earlier than MRS 3.x.

For clusters of **MRS 3.*x*** or later, see **Unlocking a User**.

Procedure

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 On MRS Manager, click **System**.

Step 3 In the **Permission** area, click **Manage User**.

Step 4 In the row of a user to be unlocked, click **Unlock User**.

Step 5 In the window that is displayed, click **OK** to unlock the user.

----End

10.13.9 Deleting a User

You can delete an MRS cluster user that is not required on MRS Manager. Deleting a user is allowed only in clusters with Kerberos authentication enabled or normal clusters with the EIP function enabled.

NOTE

If you want to create a new user with the same name as user A after deleting user A who has submitted a job on the client or MRS console, you need to delete user A's residual folders when deleting user A. Otherwise, the newly created user A may fail to submit a job.

To delete residual folders, log in to each Core node in the MRS cluster and run the following commands. In the following commands, **$user** indicates the folder named after the username.

```
cd /srv/BigData/hadoop/data1/nm/localdir/usercache/
```

```
rm -rf $user
```

The operations described in this section apply only to clusters of versions earlier than MRS 3.x.

For clusters of **MRS 3.*x*** or later, see **Deleting a User**.

Procedure

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 On MRS Manager, click **System**.

Step 3 In the **Permission** area, click **Manage User**.

Step 4 In the row that contains the user to be deleted, choose **More** > **Delete**.

Step 5 Click **OK**.

----End

10.13.10 Changing the Password of an Operation User

Scenario

Passwords of **Human-machine** system users must be regularly changed to ensure MRS cluster security. This section describes how to change passwords on MRS Manager.

If a new password policy needs to be used for the password modified by the user, follow instructions in **Modifying a Password Policy** to modify the password policy and then perform the following operations to modify the password.

NOTE

The operations described in this section apply only to clusters of versions earlier than MRS 3.x.

For clusters of **MRS 3.*x*** or later, see **Changing a User Password**.

Impact on the System

If you have downloaded a user authentication file, download it again and obtain the keytab file after modifying the password of the MRS cluster user.

Prerequisites

- You have obtained the current password policies.
- You have obtained the MRS Manager access address.
- You have obtained a cluster with Kerberos authentication enabled or a common cluster with the EIP function enabled.

Procedure

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 On MRS Manager, move the mouse cursor to in the upper right corner.

On the menu that is displayed, select **Change Password**.

Step 3 Fill in the **Old Password**, **New Password**, and **Confirm Password**. Click **OK**.

For the cluster, the default password complexity requirements are as follows:

- The password must contain 8 to 32 characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, spaces, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The password cannot be the username or the reverse username.

----End

10.13.11 Initializing the Password of a System User

Scenario

This section describes how to initialize a password on Manager if a user forgets the password or the password of a public account needs to be changed regularly. After password initialization, the user must change the password upon the first login. This operation is supported only in clusters with Kerberos authentication enabled or common clusters with the EIP function enabled.

NOTE

The operations described in this section apply only to clusters of versions earlier than MRS 3.x.

For clusters of **MRS 3.*x*** or later, see **Initializing a Password**.

Impact on the System

If you have downloaded a user authentication file, download it again and obtain the keytab file after initializing the password of the MRS cluster user.

Initializing the Password of a Human-Machine User

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 On MRS Manager, click **System**.

Step 3 In the **Permission** area, click **Manage User**.

Step 4 Locate the row that contains the user whose password is to be initialized, choose **More** > **Initialize password**, and change the password as prompted.

In the window that is displayed, enter the password of the current user and click **OK**. Then in **Initialize password**, click **OK**.

For the cluster, the default password complexity requirements are as follows:

- The password must contain 8 to 32 characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, spaces, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The password cannot be the username or the reverse username.

----End

Initializing the Password of a Machine-Machine User

Step 1 Prepare a client based on service conditions and log in to the node with the client installed.

Step 2 Run the following command to switch the user:

sudo su - omm

Step 3 Run the following command to switch to the client directory, for example, **/opt/Bigdata/client**:

cd /opt/Bigdata/client

Step 4 Run the following command to configure environment variables:

source bigdata_env

Step 5 Run the following command to log in to the console as user **kadmin/admin**:

NOTE

The default password of user **kadmin/admin** is **KAdmin@123**, which will expire upon your first login. Change the password as prompted and keep the new password secure.

kadmin -p kadmin/admin

Step 6 Run the following command to reset the password of a component running user. This operation takes effect on all servers:

cpw *Component running user name*

For example, **cpw oms/manager**.

For the cluster, the default password complexity requirements are as follows:

- The password must contain 8 to 32 characters.
- The password must contain at least three types of the following: uppercase letters, lowercase letters, digits, spaces, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
- The password cannot be the username or the reverse username.

----End

10.13.12 Downloading a User Authentication File

Scenario

When a user develops big data applications and runs them in an MRS cluster that supports Kerberos authentication, the user needs to prepare a **Machine-machine** user authentication file for accessing the MRS cluster. The keytab file in the authentication file can be used for user authentication.

This section describes how to download a **Machine-machine** user authentication file and export the keytab file on Manager. This operation is supported only in clusters with Kerberos authentication enabled or common clusters with the EIP function enabled.

NOTE

Before downloading a **Human-machine** user authentication file, change the password for the user on MRS Manager to make the initial password invalid. Otherwise, the exported keytab file cannot be used. For details, see **Changing the Password of an Operation User**.

The operations described in this section apply only to clusters of versions earlier than MRS 3.x.

For clusters of **MRS 3.*x*** or later, see **Exporting an Authentication Credential File**.

Procedure

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 On MRS Manager, click **System**.

Step 3 In the **Permission** area, click **Manage User**.

Step 4 In the row of a user for whom you want to export the keytab file, choose **More** > **Download authentication credential** to download the authentication file. After the file is automatically generated, save it to a specified path and keep it secure.

Step 5 Open the authentication file with a decompression program.

- **user.keytab** indicates a user keytab file used for user authentication.
- **krb5.conf** indicates the configuration file of the authentication server. The application connects to the authentication server according to this configuration file information when authenticating users.

----End

10.13.13 Modifying a Password Policy

Scenario

NOTICE

Because password policies are critical to the user management security, modify them based on service security requirements. Otherwise, security risks may be incurred.

This section describes how to set password and user login security rules as well as user lock rules. Password policies set on MRS Manager take effect for **Human-machine** users only, because the passwords of **Machine-machine** users are randomly generated. This operation is supported only in clusters with Kerberos authentication enabled or common clusters with the EIP function enabled.

If a new password policy needs to be used for a new user's password or the password modified by the user, perform the following operations to modify the password policy first, and then follow instructions in **Creating a User** or **Changing the Password of an Operation User**.

NOTE

The operations described in this section apply only to clusters of versions earlier than MRS 3.x.

For clusters of **MRS 3.*x*** or later, see **Configuring Password Policies**.

Procedure

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 On MRS Manager, click **System**.

Step 3 Click **Configure Password Policy**.

Step 4 Modify password policies as prompted. For parameter details, see **Table 10-52**.

Table 10-52 Password policy parameter description

Parameter	Description
Minimum Password Length	Indicates the minimum number of characters a password contains. The value ranges from 8 to 32. The default value is **8**.
Number of Character Types	Indicates the minimum number of character types a password contains. The character types include uppercase letters, lowercase letters, digits, spaces, and special characters (~`!?,.:;-_'(){}[]/<>@#$%^&*+|\=). The value can be **3** or **4**. The default value **3** indicates that the password must contain at least three types of the following characters: uppercase letters, lowercase letters, digits, special characters, and spaces.
Password Validity Period (days)	Indicates the validity period (days) of a password. The value ranges from 0 to 90. Value **0** means that the password is permanently valid. The default value is **90**.
Password Expiration Notification Days	Indicates the number of days to notify password expiration in advance. After the value is set, if the difference between the cluster time and the password expiration time is smaller than this value, the user receives password expiration notifications. When a user logs in to MRS Manager, a message is displayed, indicating that the password is about to expire and asking the user whether to change the password. The value ranges from **0** to *X* (*X* must be set to the half of the password validity period and rounded down). Value **0** indicates that no notification is sent. The default value is **5**.

Parameter	Description
Interval of Resetting Authentication Failure Count (min)	Indicates the interval (minutes) of retaining incorrect password attempts. The value ranges from 0 to 1440. Value **0** indicates that the number of incorrect password attempts are permanently retained and value **1440** indicates that the number of incorrect password attempts are retained for one day. The default value is **5**.
Number of Password Retries	Indicates the number of consecutive wrong passwords allowed before the system locks the user. The value ranges from 3 to 30. The default value is **5**.
Account Lock Duration (min)	Indicates the time period for which a user is locked when the user lockout conditions are met. The value ranges from 5 to 120. The default value is **5**.

----End

10.13.14 Configuring Cross-Cluster Mutual Trust Relationships

Scenario

If cluster A needs to access the resources of cluster B, the mutual trust relationship must be configured between these two clusters.

If no trust relationship is configured, resources of a cluster are available only for users in this cluster. MRS automatically assigns a unique **domain name** for each cluster to define the scope of resources for users.

NOTE

The operations described in this section apply only to clusters of versions earlier than MRS 3.x.

For clusters of **MRS 3.*x*** or later, see **Configuring Cross-Manager Mutual Trust Between Clusters**.

Impact on the System

- After cross-cluster mutual trust is configured, resources of a cluster become available for users in other cluster. User permission in the clusters must be regularly checked based on service and security requirements.
- After cross-cluster mutual trust is configured, the KrbServer service needs to be restarted and the cluster becomes unavailable during the restart.
- After cross-cluster mutual trust is configured, internal users **krbtgt/***Local cluster domain name***@***External cluster domain name* and **krbtgt/***External cluster domain name***@***Local cluster domain name* are added to the two

clusters. The internal users cannot be deleted. The default password is **Crossrealm@123**.

Procedure

Step 1 On the MRS management console, query all security groups of the two clusters.

- If the security groups of the two clusters are the same, go to **Step 3**.
- If the security groups of the two clusters are different, go to **Step 2**.

Step 2 On the VPC management console, add rules for each security group.

Set **Protocol** to **ANY**, **Transfer Direction** to **Inbound**,

and **Source** to **Security Group**. The source is the security group of the peer cluster.

- For cluster A, add inbound rules to the security group, set **Source** to the security groups of cluster B (the peer cluster of cluster A).
- For cluster B, add inbound rules to the security group, set **Source** to the security groups of cluster A (the peer cluster of cluster B).

NOTE

For a common cluster with Kerberos authentication disabled, perform step **Step 1** to **Step 2** to configure cross-cluster mutual trust. For a security cluster with Kerberos authentication enabled, after completing the preceding steps, proceed to the following steps for configuration.

Step 3 Log in to MRS Manager of the two clusters separately. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**. Click **Service** and check whether the **Health Status** of all components is **Good**.

- If yes, go to **Step 4**.
- If no, contact technical support personnel for troubleshooting.

Step 4 Query configuration information.

1. On MRS Manager of the two clusters, choose **Services** > **KrbServer** > **Instance**. Query the **OM IP Address** of the two KerberosServer hosts.
2. Click **Service Configuration**. Set **Type** to **All**. Choose **KerberosServer** > **Port** in the navigation tree on the left. Query the value of **kdc_ports**. The default value is **21732**.
3. Click **Realm** and query the value of **default_realm**.

Step 5 On MRS Manager of either cluster, modify the **peer_realms** parameter.

Table 10-53 Parameter description

Parameter	Description
realm_name	Domain name of the mutual-trust cluster, that is, the value of **default_realm** obtained in step **4**.

Parameter	Description
ip_port	KDC address of the peer cluster. Format: *IP address of a KerberosServer node in the peer cluster:kdc_port* The addresses of the two KerberosServer nodes are separated by a comma. For example, if the IP addresses of the KerberosServer nodes are 10.0.0.1 and 10.0.0.2 respectively, the value of this parameter is **10.0.0.1:21732,10.0.0.2:21732**.

NOTE

- To deploy trust relationships with multiple clusters, click ⊕ to add items and specify relevant parameters. To delete an item, click ⊗.
- A cluster can have trust relationships with a maximum of 16 clusters. By default, no trust relationship exists between different clusters that are trusted by a local cluster.

Step 6 Click **Save Configuration**. In the dialog box that is displayed, select **Restart the affected services or instances** and click **OK**. If you do not select **Restart the affected services or instances**, manually restart the affected services or instances.

After **Operation successful** is displayed, click **Finish**.

Step 7 Exit MRS Manager and log in to it again. If the login is successful, the configurations are valid.

Step 8 Log in to MRS Manager of the other cluster and repeat step **Step 5** to **Step 7**.

----End

Follow-up Operations

After cross-cluster mutual trust is configured, the service configuration parameters are modified on MRS Manager and the service is restarted. Therefore, you need to prepare the client configuration file again and update the client.

Scenario 1:

Cluster A and cluster B (peer cluster and mutually trusted cluster) are the same type, for example, analysis cluster or streaming cluster. Follow instructions in **Updating a Client (Versions Earlier Than 3.x)** to update the client configuration files of cluster A and B respectively.

- Update the client configuration file of cluster A.
- Update the client configuration file of cluster B.

Scenario 2:

Cluster A and cluster B (peer cluster and mutually trusted cluster) are the different type. Perform the following steps to update the configuration files.

- Update the client configuration file of cluster A to cluster B.
- Update the client configuration file of cluster B to cluster A.

- Update the client configuration file of cluster A.
- Update the client configuration file of cluster B.

Step 1 Log in to MRS Manager of cluster A.

Step 2 Click **Services**, and then **Download Client**.

Step 3 Set **Client Type** to **Only configuration files**.

Step 4 Set **Download to** to **Remote host**.

Step 5 Set **Host IP Address** to the IP address of the active Master node of cluster B, **Host Port** to 22, and **Save Path** to **/tmp**.

- If the default port **22** for logging in to cluster B using SSH is changed, set **Host Port** to a new port.
- The value of **Save Path** contains a maximum of 256 characters.

Step 6 Set **Login User** to **root**.

If another user is used, ensure that the user has permissions to read, write, and execute the save path.

Step 7 Click **OK** to generate a client file.

If the following information is displayed, the client file is saved. Click **Close**.

```
Client files downloaded to the remote host successfully.
```

If the following information is displayed, check the username, password, and security group configurations of the remote host. Ensure that the username and password are correct and an inbound rule of the SSH (22) port has been added to the security group of the remote host. And then, go to **Step 2** to download the client again.

```
Failed to connect to the server. Please check the network connection or parameter settings.
```

Step 8 Log in to the ECS of cluster B using VNC. For details, see **Instances** > **Logging In to a Windows ECS > Login Using VNC** in the *Elastic Cloud Server User Guide*

Log in to the ECS. For details, see **Login Using an SSH Key**. Set the ECS password and log in to the ECS in VNC mode.

Step 9 Run the following command to switch to the client directory, for example, **/opt/Bigdata/client**:

cd /opt/Bigdata/client

Step 10 Run the following command to update the client configuration of cluster A to cluster B:

sh refreshConfig.sh *Client installation directory Full path of the client configuration file package*

For example, run the following command:

sh refreshConfig.sh /opt/Bigdata/client /tmp/MRS_Services_Client.tar

If the following information is displayed, the configurations have been updated successfully.

```
ReFresh components client config is complete.
Succeed to refresh components client config.
```

NOTE

You can also refer to method 2 in **Updating a Client (Versions Earlier Than 3.x)** to perform operations in **Step 1** to **Step 10**.

Step 11 Repeat step **Step 1** to **Step 10** to update the client configuration file of cluster B to cluster A.

Step 12 Follow instructions in **Updating a Client (Versions Earlier Than 3.x)** to update the client configuration file of the local cluster.

- Update the client configuration file of cluster A.
- Update the client configuration file of cluster B.

----End

10.13.15 Configuring Users to Access Resources of a Trusted Cluster

Scenario

After cross-cluster mutual trust is configured, permission must be configured for users in the local cluster, so that the users can access the same resources in the peer cluster as the users in the peer cluster.

NOTE

The operations described in this section apply only to clusters of versions earlier than MRS 3.x.

For clusters of **MRS 3.*x*** or later, see **Assigning User Permissions After Cross-Cluster Mutual Trust Is Configured**.

Prerequisites

The mutual trust relationship has been configured between two clusters (clusters A and B). The clients of the clusters have been updated.

Procedure

Step 1 Log in to MRS Manager of cluster A and choose **System** > **Manage User**. Check whether cluster A has accounts that are the same as those of cluster B.

- If yes, go to **Step 2**.
- If no, go to **Step 3**.

Step 2 Click ⌄ on the left side of the username to unfold the detailed user information. Check whether the user group and role to which the user belongs meet the service requirements.

For example, user **admin** of cluster A has the permission to access and create files in the **/tmp** directory of cluster A. Then go to **Step 4**.

Step 3 Create the accounts in cluster A and bind the accounts to the user group and roles required by the services. Then go to **Step 4**.

Step 4 Choose **Service** > **HDFS** > **Instance**. Query the **OM IP Address** of **NameNode (Active)**.

Step 5 Log in to the client of cluster B.

For example, if you have updated the client on the Master2 node, log in to the Master2 node to use the client. For details, see **Using an MRS Client**.

Step 6 Run the following command to access the **/tmp** directory of cluster A.

hdfs dfs -ls hdfs://192.168.6.159:9820/tmp

In the preceding command, **192.168.6.159** is the IP address of the active NameNode of cluster A; **9820** is the default port for communication between the client and the NameNode.

Step 7 Run the following command to create a file in the **/tmp** directory of cluster A:

hdfs dfs -touchz hdfs://192.168.6.159:9820/tmp/mrstest.txt

If you can query the **mrstest.txt** file in the **/tmp** directory of cluster A, the cross-cluster mutual trust is configured successfully.

----End

10.13.16 Configuring Fine-Grained Permissions for MRS Multi-User Access to OBS

When fine-grained permission control is enabled, you can configure OBS access permissions to implement access control on directories in OBS file systems.

NOTE

This section applies only to MRS 2.*x* or earlier (excluding MRS 1.9.2).

This function enables you to control MRS users' access to OBS resources. For example, if you allow user group A to only access log files in a specified OBS file system, perform the following operations:

1. Configure an agency with OBS access permissions for an MRS cluster so that OBS can be accessed using the temporary AK/SK automatically obtained by the ECS. This prevents the AK/SK from being exposed in the configuration file.
2. Create a policy on the IAM console to allow access to log files in a specified OBS file system, and create an agency bound to the policy permission.
3. In the MRS cluster, bind the new agency to user group A so that user group A only has the permission to access log files in the specified OBS file system.

In the following scenarios, the username used for submitting jobs is an internal username so that MRS multi-user access to OBS is not supported.

- For spark-beeline, the internal username used for submitting jobs is **spark** in a security cluster and **omm** in a normal cluster.
- For the HBase shell, the internal username used for submitting jobs is **hbase** in a security cluster and **omm** in a normal cluster.
- For Presto, the internal username used for submitting jobs in the security cluster is **omm** or **hive**, and that in the normal cluster is **omm**. (Choose **Components** > **Presto** > **Service Configuration**. Change **Basic** to **All** in the

parameter type drop-down box.) Then, search for and change the value of **hive.hdfs.impersonation.enabled** to **true** to enable MRS multi-user to access OBS with fine-grained permissions.

Prerequisites

- Fine-grained permission control has been enabled. For details about permissions management, see **Creating an MRS User**.
- You have a basic knowledge of and OBS fine-grained policies.

Step 1: Configuring an Agency with OBS Access Permission for a Cluster

Step 1 Follow instructions in **Configuring a Storage-Compute Decoupled Cluster (Agency)** to configure an agency with OBS access permissions.

The agency takes effect for all users (including internal users) and user groups in the cluster. To control the permissions of users and user groups in the cluster to access OBS, perform the following operations.

----End

Step 2: Creating a Policy and an Agency on IAM

Create policies with different access permissions and bind the policies to the agency. For details, see **Creating a Policy and an Agency on IAM**.

Step 3: Configuring OBS Permission Control Mappings on the MRS Cluster Details Page

Step 1 On the MRS management console, choose **Clusters** > **Active Clusters** and click the cluster name.

Step 2 In the **Basic Information** area on the **Dashboard** tab page, click **Manage** next to **OBS Permission Control**.

Step 3 Click **Add Mapping** and set parameters according to **Table 10-54**.

Table 10-54 OBS permission control parameters

Parameter	Description
IAM Agency	Select the agency created in **Step 2**.
Type	● **User**: User-level mapping ● **Group**: User group-level mapping NOTE ● User-level mapping takes priority over user group-level mapping. If you select **Group**, you are advised to enter the primary group name in **MRS User (User Group)**. ● Do not use the same username (user group) for multiple mapping records.

Parameter	Description
MRS User (User Group)	Use commas (,) to separate multiple names of users or user groups. **NOTE** ● If OBS permission control is not configured for a user and no AK and SK are configured, the permission in **MRS_ECS_DEFAULT_AGENCY** will be used for accessing OBS. You are advised not to bind the internal user of a component to an agency. ● If you need to configure an agency for the internal user of a component when submitting a job in the following scenarios, the requirements are as follows: – To control permissions on spark-beeline operations, set the username to **spark** for a security cluster and **omm** for a normal cluster. – To control permissions on HBase shell operations, set the username to **hbase** for a security cluster and **omm** for a normal cluster. – To control permissions on Presto, set the username to **omm**, **hive**, and the username used for logging in to the client for a security cluster and **omm** and the username used for logging in to the client for a normal cluster. – If you want to use Hive to create tables in beeline mode, set the username to the internal user **hive**.

Step 4 Click **OK**.

Step 5 Select **I agree to authorize the trust relationships between MRS Users (Groups) and IAM agencies**, and click **OK**. The mapping between the MRS user and OBS permission is added.

If ✓ appears next to **OBS Permission Control** on the **Dashboard** tab page or the mapping table has been updated for OBS permission control, the mapping takes effect. It takes about 1 minute to for the mapping to take effect.

In the **Operation** column of the mapping list, you can edit or delete the added mapping.

NOTE

- If OBS permission control is not configured for a user and no AK and SK are configured, the permissions owned by the agency configured for the cluster in the **Object Storage Service (OBS)** project will be used to access OBS.
- Regardless of whether OBS permission control is configured, AK/SK permission is used for accessing OBS once it is configured.
- Security Administrator permission is required to modify, create, or delete a mapping.
- To enable mapping changes to take effect in spark-line, hive beeline and Presto respectively, you need to restart Spark, exit beeline and enter again, and restart Presto respectively.

----End

Component Access to OBS When OBS Permission Control Is Enabled

Step 1 Log in to any node in a cluster as user **root** using the password set during cluster creation.

Step 2 Set environment variables (The default client installation path for MRS 3.x or later is /opt/Bigdata/client, and for versions earlier than MRS 3.x is /opt/client. Configure the path based on site requirements.).

source /opt/Bigdata/client/bigdata_env

Step 3 If the Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If the Kerberos authentication is disabled for the current cluster, skip this step:

kinit *MRS cluster user*

Example: **kinit admin**

Step 4 If the Kerberos authentication is disabled for the current cluster, run the following commands to log in. Note that you should create a user that belongs to the **supergroup** group by referring to **Creating a User** and replace *XXXX* with the username:

mkdir /home/XXXX

chown XXXX /home/XXXX

su - XXXX

Step 5 Access OBS. You do not need to configure the AK, SK, and endpoint. The OBS path format is **obs://buck_name/*XXX***.

Example: **hadoop fs -ls "obs://obs-example/job/hadoop-mapreduce-examples-3.1.2.jar"**

NOTE

- If you want to use **hadoop fs** to delete files on OBS, use **hadoop fs -rm -skipTrash** to delete the files.
- If data import is not involved when a table is created using spark-sql and spark-beeline, OBS will not be accessed. That is, if you create a table in an OBS directory on which you do not have permission, the **CREATE TABLE** operation will still be successful, but the error message "**403 AccessDeniedException**" is displayed when you insert data.

----End

Creating a Policy and an Agency on IAM

Step 1 Create a policy on IAM.

1. Log in to the IAM console.
2. Choose **Permissions**. On the displayed page, click **Create Custom Policy**.
3. Set parameters according to **Table 10-55**. Obtain the customized OBS policy samples that are frequently used by referring to .

Table 10-55 Policy parameters

Parameter	Description
Policy Name	Only letters, digits, spaces, and special characters (-_.,) are allowed.
Scope	Select **Global services**, because OBS is a global service.
Policy View	Select **Visual editor**.

Parameter	Description
Policy Content	1. **Allow**: Select **Allow**. 2. **Select service**: Select **Object Storage Service (OBS)**. 3. **Select action**: Select **WriteOnly**, **ReadOnly**, and **ListOnly**. 4. **Specific resources**: a. Set **object** to **Specify resource path**, click **Add Resource Path**, and enter *obs_bucket_name*/**tmp/** and *obs_bucket_name*/**tmp/***. The **/tmp** directory is used as an example. If you need to add permissions for other directories, perform the following steps to add the directories and resource paths of all objects in the directories. b. Set **bucket** to **Specify resource path**, click **Add Resource Path**, and enter *obs_bucket_name*. 5. (Optional) Add request condition, which does not need to be added currently.
Description	(Optional) Brief description about the policy.

NOTE

If the data write operation of each component is implemented in **rename** mode, the permission to delete objects must be configured when data is written.

4. Click **OK** to save the policy.

Step 2 Create an agency on IAM.

1. Log in to the IAM console.
2. Choose **Agencies**. On the displayed page, click **Create Agency**.
3. Set parameters according to **Table 10-56**.

Table 10-56 Agency parameters

Parameter	Description
Agency Name	Only letters, digits, spaces, and special characters (-_.,) are allowed.

Parameter	Description
Agency Type	Select **Common account**.
Delegated Account	Enter your cloud account, that is, the account you register using your mobile phone number. It cannot be a federated user or an IAM user created using your cloud account.
Validity Period	Set this parameter as required.
Description	(Optional) Brief description about the agency.
Permissions	1. In the **Project [Region]** column, locate the row where **OBS** is, click **Attach Policy**. 2. Select the policy created in **Step 1** to display it in **Selected Policies**. 3. Click **OK**.

4. Click **OK** to save the agency.

NOTE

If you modify an agency and policies bound to it after using the agency to access OBS, the modification will take effect within 15 minutes.

----End

10.14 Patch Operation Guide

10.14.1 Patch Operation Guide for Versions Earlier than MRS 1.7.0

If you obtain patch information from the following sources, upgrade the patch according to actual requirements.

- You obtain information about the patch released by MRS from a message pushed by the message center service.
- You obtain information about the patch by accessing the cluster and viewing patch information.

Preparing for Patch Installation

- Follow instructions in **Performing a Health Check** to check cluster status. If the cluster health status is normal, install a patch.
- You have uploaded the cluster patch package to the server. For details, see **Uploading the Patch Package**.

- You need to confirm the target patch to be installed according to the patch information in the patch content.

Uploading the Patch Package

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 Choose **System** > **Manage Patch**. The **Manage Patch** page is displayed.

Step 3 Click **Upload Patch** and set the following parameters.

- **Patch File Path**: Folder created in the OBS file system where the patch package is stored, for example, **MRS_1.6.2/MRS_1_6_2_11.tar.gz**
- **Bucket**: Name of the OBS file system where the patch package is stored, for example, **mrs_patch**

NOTE

You can obtain the file system name and patch file path on the **Patch Information** tab page. The value of the **Patch Path** is in the following format: [File system name]/[Patch file path].

- **AK**: For details, see .
- **SK**: For details, see .

Step 4 Click **OK** to upload the patch.

----End

Installing a Patch

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 Choose **System** > **Manage Patch**. The **Manage Patch** page is displayed.

Step 3 In the **Operation** column, click **Install**.

Step 4 In the displayed dialog box, click **OK** to install the patch.

Step 5 After the patch is installed, you can view the installation status in the progress bar. If the installation fails, contact the MRS cluster administrator.

NOTE

For the isolated host nodes in the cluster, follow instructions in **Restoring Patches for the Isolated Hosts** to restore the patch.

----End

Uninstalling a Patch

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 Choose **System** > **Manage Patch**. The **Manage Patch** page is displayed.

Step 3 In the **Operation** column, click **Uninstall**.

NOTE

For the isolated host nodes in the cluster, follow instructions in **Restoring Patches for the Isolated Hosts** to restore the patch.

----End

10.14.2 Patch Operation Guide for Versions

If you obtain patch information from the following sources, upgrade the patch according to actual requirements.

- You obtain information about the patch released by MRS from a message pushed by the message center service.
- You obtain information about the patch by accessing the cluster and viewing patch information.

Preparing for Patch Installation

- Follow instructions in **Performing a Health Check** to check cluster status. If the cluster health status is normal, install a patch.
- You need to confirm the target patch to be installed according to the patch information in the patch content.

Installing a Patch

Step 1 Log in to the MRS management console.

Step 2 Choose **Clusters** > **Active Clusters** and click the name of the cluster to be queried to enter the page displaying the cluster's basic information.

Step 3 On the **Patch Information** page, click **Install** in the **Operation** column to install the target patch.

NOTE

- For details about rolling patch operations, see **Supporting Rolling Patches**.
- For the isolated host nodes in the cluster, follow instructions in **Restoring Patches for the Isolated Hosts** to restore the patch.

----End

Uninstalling a Patch

Step 1 Log in to the MRS management console.

Step 2 Choose **Clusters** > **Active Clusters** and click the name of the cluster to be queried to enter the page displaying the cluster's basic information.

Step 3 On the **Patch Information** page, click **Uninstall** in the **Operation** column to uninstall the target patch.

NOTE

- For details about rolling patch operations, see **Supporting Rolling Patches**.
- For the isolated host nodes in the cluster, follow instructions in **Restoring Patches for the Isolated Hosts** to restore the patch.

----End

10.14.3 Supporting Rolling Patches

The rolling patch function indicates that patches are installed or uninstalled for one or more services in a cluster by performing a rolling service restart (restarting services or instances in batches), without interrupting the services or within a minimized service interruption interval. Services in a cluster are divided into the following three types based on whether they support rolling patch:

- Services supporting rolling patch installation or uninstallation: All businesses or part of them (varying depending on different services) of the services are not interrupted during patch installation or uninstallation.
- Services not supporting rolling patch installation or uninstallation: Businesses of the services are interrupted during patch installation or uninstallation.
- Services with some roles supporting rolling patch installation or uninstallation: Some businesses of the services are not interrupted during patch installation or uninstallation.

Table 10-57 provides services and instances that support or do not support rolling restart in the MRS cluster.

Table 10-57 Services and instances that support or do not support rolling restart

Service	Instance	Whether to Support Rolling Restart
HDFS	NameNode	Yes
	ZKFC	
	JournalNode	
	HttpFS	
	DataNode	
Yarn	ResourceManager	Yes
	NodeManager	
Hive	MetaStore	Yes
	WebHCat	
	HiveServer	
MapReduce	JobHistoryServer	Yes
HBase	HMaster	Yes

Service	Instance	Whether to Support Rolling Restart
	RegionServer	
	ThriftServer	
	RESTServer	
Spark	JobHistory	Yes
	JDBCServer	
	SparkResource	No
Hue	Hue	No
Tez	TezUI	No
Loader	Sqoop	No
ZooKeeper	QuorumPeer	Yes
Kafka	Broker	Yes
	MirrorMaker	No
Flume	Flume	Yes
	MonitorServer	
Storm	Nimbus	Yes
	UI	
	Supervisor	
	LogViewer	

Installing a Patch

Step 1 Log in to the MRS management console.

Step 2 Choose **Clusters** > **Active Clusters** and click the name of the cluster to be queried to enter the page displaying the cluster's basic information.

Step 3 On the **Patch Information** page, click **Install** in the **Operation** column.

Step 4 On the **Warning** page, enable or disable **Rolling Patch**.

NOTE

- Enabling the rolling patch installation function: Services are not stopped before patch installation, and rolling service restart is performed after the patch installation. This minimizes the impact on cluster services but takes more time than common patch installation.
- Disabling the rolling patch uninstallation function: All services are stopped before patch uninstallation, and all services are restarted after the patch uninstallation. This temporarily interrupts the cluster and the services but takes less time than rolling patch uninstallation.
- The rolling patch installation function is not available in clusters with less than two Master nodes and three Core nodes.

Step 5 Click **OK** to install the target patch.

Step 6 View the patch installation progress.

1. Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.
2. Choose **System** > **Manage Patch**. On the **Manage Patch** page, you can view the patch installation progress.

NOTE

For the isolated host nodes in the cluster, follow instructions in **Restoring Patches for the Isolated Hosts** to restore the patch.

----End

Uninstalling a Patch

Step 1 Log in to the MRS management console.

Step 2 Choose **Clusters** > **Active Clusters** and click the name of the cluster to be queried to enter the page displaying the cluster's basic information.

Step 3 On the **Patch Information** page, click **Uninstall** in the **Operation** column.

Step 4 On the **Warning** page, enable or disable **Rolling Patch**.

NOTE

- Enabling the rolling patch uninstallation function: Services are not stopped before patch uninstallation, and rolling service restart is performed after the patch uninstallation. This minimizes the impact on cluster services but takes more time than common patch uninstallation.
- Disabling the rolling patch uninstallation function: All services are stopped before patch uninstallation, and all services are restarted after the patch uninstallation. This temporarily interrupts the cluster and the services but takes less time than rolling patch uninstallation.
- The rolling patch uninstallation function is not available in clusters with less than two Master nodes and three Core nodes.

Step 5 Click **OK** to uninstall the target patch.

Step 6 View the patch uninstallation progress.

1. Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

2. Choose **System** > **Manage Patch**. On the **Manage Patch** page, you can view the patch uninstallation progress.

NOTE

For the isolated host nodes in the cluster, follow instructions in **Restoring Patches for the Isolated Hosts** to restore the patch.

----End

10.15 Restoring Patches for the Isolated Hosts

If some hosts are isolated in a cluster, perform the following operations to restore patches for these isolated hosts after patch installation on other hosts in the cluster. After patch restoration, versions of the isolated host nodes are consistent with those are not isolated.

Step 1 Access MRS Manager. For details, see **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 2 Choose **System** > **Manage Patch**. The **Manage Patch** page is displayed.

Step 3 In the **Operation** column, click **View Details**.

Step 4 On the patch details page, select host nodes whose **Status** is **Isolated**.

Step 5 Click **Select and Restore** to restore the isolated host nodes.

----End

10.16 Rolling Restart

After modifying the configuration items of a big data component, you need to restart the corresponding service to make new configurations take effect. If you use a normal restart mode, all services or instances are restarted concurrently, which may cause service interruption. To ensure that services are not affected during service restart, you can restart services or instances in batches by rolling restart. For instances in active/standby mode, a standby instance is restarted first and then an active instance is restarted. Rolling restart takes longer than normal restart.

Table 10-58 provides services and instances that support or do not support rolling restart in the MRS cluster.

Table 10-58 Services and instances that support or do not support rolling restart

Service	Instance	Whether to Support Rolling Restart
HDFS	NameNode	Yes
	ZKFC	
	JournalNode	

Service	Instance	Whether to Support Rolling Restart
	HttpFS	
	DataNode	
Yarn	ResourceManager	Yes
	NodeManager	
Hive	MetaStore	Yes
	WebHCat	
	HiveServer	
MapReduce	JobHistoryServer	Yes
HBase	HMaster	Yes
	RegionServer	
	ThriftServer	
	RESTServer	
Spark	JobHistory	Yes
	JDBCServer	
	SparkResource	No
Hue	Hue	No
Tez	TezUI	No
Loader	Sqoop	No
ZooKeeper	Quorumpeer	Yes
Kafka	Broker	Yes
	MirrorMaker	No
Flume	Flume	Yes
	MonitorServer	
Storm	Nimbus	Yes
	UI	
	Supervisor	
	Logviewer	

Restrictions

- Perform a rolling restart during off-peak hours.
  - Otherwise, a rolling restart failure may occur. For example, if the throughput of Kafka is high (over 100 MB/s) during the Kafka rolling restart, the Kafka rolling restart may fail.
  - For example, if the requests per second of each RegionServer on the native interface exceed 10,000 during the HBase rolling restart, you need to increase the number of handles to prevent a RegionServer restart failure caused by heavy loads during the restart.
- Before the restart, check the number of current requests of HBase. If requests of each RegionServer on the native interface exceed 10,000, increase the number of handles to prevent a failure.
- If the number of Core nodes in a cluster is less than six, services may be affected for a short period of time.
- Preferentially perform a rolling instance or service restart and select **Only restart instances whose configurations have expired**.

Performing a Rolling Service Restart

Step 1 On MRS Manager, click **Services** and select a service for which you want to perform a rolling restart.

Step 2 On the **Service Status** tab page, click **More** and select **Perform Rolling Service Restart**.

Step 3 After you enter the administrator password, the **Perform Rolling Service Restart** page is displayed. Select **Only restart instances whose configurations have expired** and click **OK** to perform rolling restart for the service.

Step 4 After the rolling restart task is complete, click **Finish**.

----End

Performing a Rolling Instance Restart

Step 1 On MRS Manager, click **Services** and select a service for which you want to perform a rolling restart.

Step 2 On the **Instance** tab page, select the instance to be restarted. Click **More** and select **Perform Rolling Instance Restart**.

Step 3 After you enter the password, the **Perform Rolling Instance Restart** page is displayed. Select **Only restart instances whose configurations have expired** and click **OK** to perform rolling restart for the instance.

Step 4 After the rolling restart task is complete, click **Finish**.

----End

Perform a Rolling Cluster Restart

Step 1 On MRS Manager, click **Services**. The **Services** page is displayed.

Step 2 Click **More** and select **Perform Rolling Cluster Restart**.

Step 3 After you enter the password, the **Perform Rolling Cluster Restart** page is displayed. Select **Only restart instances whose configurations have expired** and click **OK** to perform rolling restart for the cluster.

Step 4 After the rolling restart task is complete, click **Finish**.

----End

Rolling Restart Parameter Description

Table 10-59 describes rolling restart parameters.

Table 10-59 Rolling restart parameter description

Parameter	Description
Only restart instances whose configurations have expired	Specifies whether to restart only the modified instances in a cluster.
Data Node Instances to Be Batch Restarted	Specifies the number of instances that are restarted in each batch when the batch rolling restart strategy is used. The default value is **1**. The value ranges from 1 to 20. This parameter is valid only for data nodes.
Batch Interval	Specifies the interval between two batches of instances for rolling restart. The default value is **0**. The value ranges from 0 to 2147483647. The unit is second. Note: Setting the batch interval parameter can increase the stability of the big data component process during the rolling restart. You are advised to set this parameter to a non-default value, for example, 10.
Batch Fault Tolerance Threshold	Specifies the tolerance times when the rolling restart of instances fails to be executed in batches. The default value is **0**, which indicates that the rolling restart task ends after any batch of instances fails to be restarted. The value ranges from 0 to 214748364.

Procedure in a Typical Scenario

Step 1 On MRS Manager, click **Services** and select HBase. The HBase service page is displayed.

Step 2 Click the **Service Configuration** tab, and modify an HBase parameter. After the following dialog box is displayed, click **OK** to save the configurations.

NOTE

Do not select **Restart the affected services or instances**. This option indicates a normal restart. If you select this option, all services or instances will be restarted, which may cause service interruption.

Step 3 After saving the configurations, click **Finish**.

Step 4 Click the **Service Status** tab.

Step 5 On the **Service Status** tab page, click **More** and select **Perform Rolling Service Restart**.

Step 6 After you enter the password, the **Perform Rolling Service Restart** page is displayed. Select **Only restart instances whose configurations have expired** and click **OK** to perform rolling restart for the service.

Step 7 After the rolling restart task is complete, click **Finish**.

----End

11 Security Description

11.1 Security Configuration Suggestions for Clusters with Kerberos Authentication Disabled

The Hadoop community version provides two authentication modes: Kerberos authentication (security mode) and Simple authentication (normal mode). When creating a cluster, you can choose to enable or disable Kerberos authentication.

Clusters in security mode use the Kerberos protocol for security authentication.

In normal mode, MRS cluster components use a native open source authentication mechanism, which is typically Simple authentication. If Simple authentication is used, authentication is automatically performed by a client user (for example, user **root**) by default when a client connects to a server. The authentication is imperceptible to the MRS cluster administrator or service user. In addition, when being executed, the client may even pretend to be any user (including **superuser**) by injecting **UserGroupInformation**. Cluster resource management and data control APIs are not authenticated on the server and are easily exploited and attacked by hackers.

Therefore, in normal mode, network access permissions must be strictly controlled to ensure cluster security. You are advised to perform the following operations to ensure cluster security.

- Deploy service applications on ECSs in the same VPC and subnet and avoid accessing MRS clusters through an external network.
- Configure security group rules to strictly control the access scope. Do not configure access rules that allow **Any** or **0.0.0.0** for the inbound direction of MRS cluster ports.
- If you want to access the native pages of the components in the cluster from the external, follow instructions in **Creating an SSH Channel for Connecting to an MRS Cluster and Configuring the Browser** for configuration.

11.2 Security Authentication Principles and Mechanisms

Function

For clusters in security mode with Kerberos authentication enabled, security authentication is required during application development.

Kerberos, named after the ferocious three-headed guard dog of Hades from Greek mythology, is now used to a concept in authentication. The Kerberos protocol adopts a client–server model and cryptographic algorithms such as AES (Advanced Encryption Standard). It provides mutual authentication, that is, both the client and the server can verify each other's identity. Kerberos is used to prevent interception and replay attacks and protect data integrity. It is a system that manages keys by using a symmetric key mechanism.

Architecture

Kerberos architecture is shown in **Figure 11-1** and module description in **Table 11-1**.

Figure 11-1 Kerberos architecture

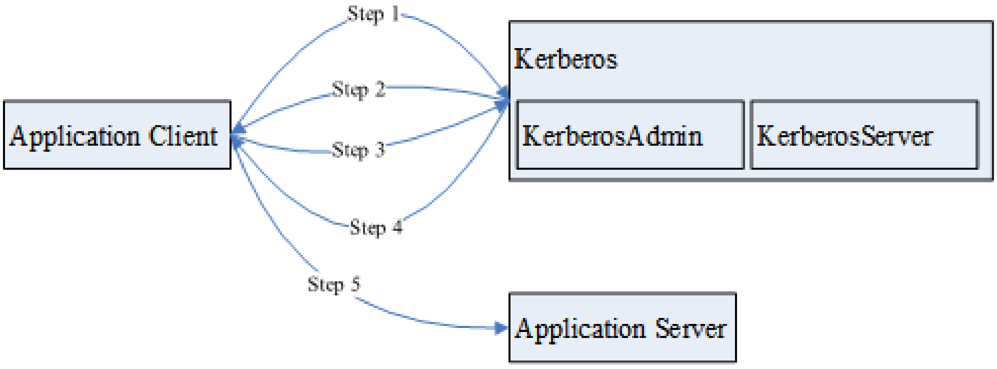


Table 11-1 Module description

Module	Description
Application Client	An application client, which is usually an application that submits tasks or jobs
Application Server	An application server, which is usually an application that an application client accesses
Kerberos	A service that provides security authentication

Module	Description
KerberosAdmin	A process that provides authentication user management
KerberosServer	A process that provides authentication ticket distribution

The process and principle are described as follows:

An application client can be a service in a cluster or a secondary development application of the customer. An application client can submit tasks or jobs to an application service.

1. Before submitting a task or job, the application client needs to apply for a ticket granting ticket (TGT) from the Kerberos service to establish a secure session with the Kerberos server.
2. After receiving the TGT request, the Kerberos service resolves parameters in the request to generate a TGT, and uses the key of the username specified by the client to encrypt the response.
3. After receiving the TGT response, the application client (based on the underlying RPC) resolves the response and obtains the TGT, and then applies for a server ticket (ST) of the application server from the Kerberos service.
4. After receiving the ST request, the Kerberos service verifies the TGT validity in the request and generates an ST of the application service, and then uses the application service key to encrypt the response.
5. After receiving the ST response, the application client packages the ST into a request and sends the request to the application server.
6. After receiving the request, the application server uses its local application service key to resolve the ST. After successful verification, the request becomes valid.

Basic Concepts

The following concepts can help users learn the Kerberos architecture quickly and understand the Kerberos service better. The following uses security authentication for HDFS as an example.

TGT

A TGT is generated by the Kerberos service and used to establish a secure session between an application and the Kerberos server. The validity period of a TGT is 24 hours. After 24 hours, the TGT expires automatically.

The following describes how to apply for a TGT (HDFS is used as an example):

1. Obtain a TGT through an API provided by HDFS.

```
/**
 * login Kerberos to get TGT, if the cluster is in security mode
 * @throws IOException if login is failed
 */
 private void login() throws IOException {
 // not security mode, just return
   if (! "kerberos".equalsIgnoreCase(conf.get("hadoop.security.authentication"))) {
       return;
   }

   //security mode
   System.setProperty("java.security.krb5.conf", PATH_TO_KRB5_CONF);

   UserGroupInformation.setConfiguration(conf);
   UserGroupInformation.loginUserFromKeytab(PRNCIPAL_NAME, PATH_TO_KEYTAB);
 }
```

2. Run shell commands on the client in kinit mode.

ST

An ST is generated by the Kerberos service and used to establish a secure session between an application and application service. An ST is valid only once.

In FusionInsight products, the generation of an ST is based on the Hadoop-RPC communication. The underlying RPC submits a request to the Kerberos server and the Kerberos server generates an ST.

Sample Authentication Code

```
package com.xxx.bigdata.hdfs.examples;

import java.io.IOException;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.FileStatus;
import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.security.UserGroupInformation;

public class KerberosTest {
    private static String PATH_TO_HDFS_SITE_XML = KerberosTest.class.getClassLoader().getResource("hdfs-site.xml")
            .getPath();
    private static String PATH_TO_CORE_SITE_XML = KerberosTest.class.getClassLoader().getResource("core-site.xml")
            .getPath();
    private static String PATH_TO_KEYTAB =
KerberosTest.class.getClassLoader().getResource("user.keytab").getPath();
    private static String PATH_TO_KRB5_CONF =
KerberosTest.class.getClassLoader().getResource("krb5.conf").getPath();
    private static String PRNCIPAL_NAME = "develop";
    private FileSystem fs;
    private Configuration conf;

    /**
     * initialize Configuration
     */
    private void initConf() {
        conf = new Configuration();

        // add configuration files
        conf.addResource(new Path(PATH_TO_HDFS_SITE_XML));
        conf.addResource(new Path(PATH_TO_CORE_SITE_XML));
    }

    /**
     * login Kerberos to get TGT, if the cluster is in security mode
     * @throws IOException if login is failed
     */
    private void login() throws IOException {
        // not security mode, just return
        if (! "kerberos".equalsIgnoreCase(conf.get("hadoop.security.authentication"))) {
            return;
        }
        //security mode
        System.setProperty("java.security.krb5.conf", PATH_TO_KRB5_CONF);

        UserGroupInformation.setConfiguration(conf);
        UserGroupInformation.loginUserFromKeytab(PRNCIPAL_NAME, PATH_TO_KEYTAB);
    }

    /**
     * initialize FileSystem, and get ST from Kerberos
     * @throws IOException
     */
    private void initFileSystem() throws IOException {
        fs = FileSystem.get(conf);
    }

    /**
     * An example to access the HDFS
     * @throws IOException
     */
    private void doSth() throws IOException {
        Path path = new Path("/tmp");
        FileStatus fStatus = fs.getFileStatus(path);
        System.out.println("Status of " + path + " is " + fStatus);
        //other thing
    }


    public static void main(String[] args) throws Exception {
        KerberosTest test = new KerberosTest();
        test.initConf();
        test.login();
        test.initFileSystem();
        test.doSth();
    }
}
```

NOTE

1. During Kerberos authentication, you need to configure the file parameters required for configuring the Kerberos authentication, including the keytab path, Kerberos authentication username, and the **krb5.conf** configuration file of the client for Kerberos authentication.
2. Method **login()** indicates calling the Hadoop API to perform Kerberos authentication and generating a TGT.
3. Method **doSth** indicates calling the Hadoop API to access the file system. In this situation, the underlying RPC automatically carries the TGT to Kerberos for verification and then an ST is generated.

12 High-Risk Operations

Forbidden Operations

Table 12-1 lists forbidden operations during the routine cluster operation and maintenance process.

Table 12-1 Forbidden operations

Item	Risk
Delete ZooKeeper data directories.	ClickHouse, HDFS, Yarn, HBase, and Hive depend on ZooKeeper, which stores metadata. This operation has adverse impact on normal operating of related components.
Performing switchover frequently between active and standby JDBCServer nodes	This operation may interrupt services.
Delete Phoenix system tables and data (SYSTEM.CATALOG, SYSTEM.STATS, SYSTEM.SEQUENCE, and SYSTEM. FUNCTION).	This operation will cause service operation failures.
Manually modify data in the Hive metabase (hivemeta database).	This operation may cause Hive data parse errors. As a result, Hive cannot provide services.
Do not manually perform **INSERT** or **UPDATE** operations on Hive metadata tables.	This operation may cause Hive data parse errors. As a result, Hive cannot provide services.
Change permission on the Hive private file directory **hdfs:///tmp/hive-scratch**.	This operation may cause unavailable Hive services.
Modify **broker.id** in the Kafka configuration file.	This operation may cause invalid node data.

Item	Risk
Modify the host names of nodes.	Instances and upper-layer components on the host cannot provide services properly. The fault cannot be rectified.
Reinstall the OS of a node.	This operation will cause MRS cluster exceptions, leaving MRS clusters in abnormal status.
Use private images.	This operation will cause MRS cluster exceptions, leaving MRS clusters in abnormal status.

The following tables list the high-risk operations during the operation and maintenance of each component.

High-Risk Operations on a Cluster

Table 12-2 High-risk operations on a cluster

Operation	Risk	Severity	Workaround	Check Item
Modify the file directory or file permissions of user **omm** without permission.	This operation will lead to MRS service unavailability.	▲ ▲ ▲ ▲ ▲	Do not perform this operation.	Check whether the MRS cluster service is available.
Bind an EIP.	This operation exposes the Master node hosting MRS Manager of the cluster to the public network, increasing the risk of network attacks from the Internet.	▲ ▲ ▲ ▲ ▲	Ensure that the bound EIP is a trusted public IP address.	None

Operation	Risk	Severity	Workaround	Check Item
Enable security group rules for port 22 of a cluster.	This operation increases the risk of exploiting vulnerability of port 22.	▲ ▲ ▲ ▲ ▲	Configure a security group rule for port 22 to allow only trusted IP addresses to access the port. You are not advised to configure the inbound rule to allow 0.0.0.0 to access the port.	None
Delete a cluster or cluster data.	Data will get lost.	▲ ▲ ▲ ▲ ▲	Before deleting the data, confirm the necessity of the operation and ensure that the data has been backed up.	None
Scale in a cluster.	Data will get lost.	▲ ▲ ▲ ▲ ▲	Before scaling in the cluster, confirm the necessity of the operation and ensure that the data has been backed up.	None
Detach or format a data disk.	Data will get lost.	▲ ▲ ▲ ▲ ▲	Before performing this operation, confirm the necessity of the operation and ensure that the data has been backed up.	None

Manager High-Risk Operations

Table 12-3 Manager high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Change the OMS password.	This operation will restart all processes of OMSServer, which has adverse impact on cluster maintenance and management.	▲▲▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check whether there are uncleared alarms and whether the cluster management and maintenance are normal.
Import the certificate.	This operation will restart OMS processes and the entire cluster, which has adverse impact on cluster maintenance and management and services.	▲▲▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Perform an upgrade.	This operation will restart Manager and the entire cluster, affecting management, maintenance, and services of the cluster. Strictly manage the user who is eligible to assign the cluster management permission to prevent security risks.	▲▲▲	Ensure that there is no other maintenance and management operations when the operation is performed.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.

Operation	Risk	Severity	Workaround	Check Item
Restore the OMS.	This operation will restart Manager and the entire cluster, affecting management, maintenance, and services of the cluster.	▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Change an IP address.	This operation will restart Manager and the entire cluster, affecting management, maintenance, and services of the cluster.	▲ ▲ ▲	Ensure that there is no other maintenance and management operations when the operation is performed and that the new IP address is correct.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Change log levels.	If the log level is changed to **DEBUG**, Manager responds slowly.	▲ ▲	Before the modification, confirm the necessity of the operation and change it back to the default log level in time.	None

Operation	Risk	Severity	Workaround	Check Item
Replace a control node.	This operation will interrupt services deployed on the node. If the node is a management node, the operation will restart all OMS processes, affecting the cluster management and maintenance.	▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Replace a management node.	This operation will interrupt services deployed on the node. As a result, OMS processes will be restarted, affecting the cluster management and maintenance.	▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Restart the upper-layer service at the same time during the restart of a lower-layer service.	This operation will interrupt the upper-layer service, affecting the management, maintenance, and services of the cluster.	▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.

Operation	Risk	Severity	Workaround	Check Item
Modify the OLDAP port.	This operation will restart the LdapServer and Kerberos services and all associated services, affecting service running.	▲ ▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	None
Delete the **supergroup** group.	Deleting the **supergroup** group decreases user rights, affecting service access.	▲ ▲ ▲ ▲ ▲	Before the change, confirm the rights to be added. Ensure that the required rights have been added before deleting the **supergroup** rights to which the user is bound, ensuring service continuity.	None
Restart a service.	Services will be interrupted during the restart. If you select and restart the upper-layer service, the upper-layer services that depend on the service will be interrupted.	▲ ▲ ▲	Confirm the necessity of restarting the system before the operation.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.

Operation	Risk	Severity	Workaround	Check Item
Change the default SSH port No.	After the default port (22) is changed, functions such as cluster creation, service/instance adding, host adding, and host reinstallation cannot be used, and results of cluster health check items for node mutual trust, **omm**/**ommdba** user password expiration, and others are incorrect.	▲ ▲ ▲	Before performing this operation, restore the SSH port to the default value.	None

ClickHouse High-Risk Operations

Table 12-4 ClickHouse high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Delete data directories.	This operation may cause service information loss.	▲ ▲ ▲	Do not delete data directories manually.	Check whether data directories are normal.
Remove ClickHouseServer instances.	The ClickHouseServer instance nodes in the same shard must be removed in at the same time. Otherwise, the topology information of the logical cluster is incorrect. Before performing this operation, check the database and data table information of each node in the logical cluster and perform scale-in pre-analysis to ensure that data is successfully migrated during the scale-in process to prevent data loss	▲ ▲ ▲ ▲ ▲	Before scale-in, collect information in advance to learn the status of the ClickHouse logical cluster and instance nodes.	Check the ClickHouse logical cluster topology information, database and data table information in each ClickHouseServer instance, and data volume.

Operation	Risk	Severity	Workaround	Check Item
Add ClickHouseServer instances.	When performing this operation, you must check whether a database or data table with the same name as that on the old node needs to be created on the new node. Otherwise, subsequent data migration, data balancing, scale-in, and decommissioning will fail.	▲ ▲ ▲ ▲ ▲	Before scale-out, confirm the function and purpose of new ClickHouseServer instances and determine whether to create related databases and data tables.	Check the ClickHouse logical cluster topology information, database and data table information in each ClickHouseServer instance, and data volume.
Decommission ClickHouseServer instances.	The ClickHouseServer instance nodes in the same shard must be decommissioned in at the same time. Otherwise, the topology information of the logical cluster is incorrect. Before performing this operation, check the database and data table information of each node in the logical cluster and perform decommissioning pre-analysis to ensure that data is successfully migrated during the decommissioning process to prevent data loss	▲ ▲ ▲ ▲ ▲	Before decommissioning, collect information in advance to learn the status of the ClickHouse logical cluster and instance nodes.	Check the ClickHouse logical cluster topology information, database and data table information in each ClickHouseServer instance, and data volume.
Recommission ClickHouseServer instances.	When performing this operation, you must select all nodes in the original shard. Otherwise, the topology information of the logical cluster is incorrect.	▲ ▲ ▲ ▲ ▲	Before recommissioning, you need to confirm the home information about the shards of the node to be recommissioned.	Check the ClickHouse logical cluster topology information.

Operation	Risk	Severity	Workaround	Check Item
Modify data directory content (file and folder creation).	This operation may cause the ClickHouse instance of the node faults.	▲ ▲ ▲	Do not create or modify files or folders in the data directories manually.	Check whether data directories are normal.
Start or stop basic components independently.	This operation has adverse impact on the basic functions of some services. As a result, service failures occur.	▲ ▲ ▲	Do not start or stop ZooKeeper, Kerberos, and LDAP basic components independently. Select related services when performing this operation.	Check whether the service status is normal.
Restart or stop services.	This operation may interrupt services.	▲ ▲	Restart or stop services when necessary.	Check whether the service is running properly.

DBService High-Risk Operations

Table 12-5 DBService high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Change the DBService password.	The services need to be restarted for the password change to take effect. The services are unavailable during the restart.	▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check whether there are uncleared alarms and whether the cluster management and maintenance are normal.
Restore DBService data.	After the data is restored, the data generated after the data backup and before the data restoration is lost. After the data is restored, the configuration of the components that depend on DBService may expire and these components need to be restarted.	▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check whether there are uncleared alarms and whether the cluster management and maintenance are normal.

Operation	Risk	Severity	Workaround	Check Item
Perform active/standby DBService switchover.	During the DBServer switchover, DBService is unavailable.	▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	None
Change the DBService floating IP address.	The DBService needs to be restarted for the change to take effect. The DBService is unavailable during the restart. If the floating IP address has been used, the configuration will fail, and the DBService will fail to be started.	▲ ▲ ▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.

Flink High-Risk Operations

Table 12-6 Flink high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Change log levels.	If the log level is modified to DEBUG, the task running performance is affected.	▲ ▲	Before the modification, confirm the necessity of the operation and change it back to the default log level in time.	None

Operation	Risk	Severity	Workaround	Check Item
Modify file permissions.	Tasks may fail.	▲ ▲ ▲	Confirm the necessity of the operation before the modification.	Check whether related service operations are normal.

Flume High-Risk Operations

Table 12-7 Flume high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify the Flume instance start parameter **GC_OPTS**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Change the default value of **dfs.replication** from **3** to **1**.	This operation will have the following impacts: 1. The storage reliability deteriorates. If the disk becomes faulty, data will be lost. 2. NameNode fails to be restarted, and the HDFS service is unavailable.	▲ ▲ ▲ ▲	When modifying related configuration items, check the parameter description carefully. Ensure that there are more than two replicas for data storage.	Check whether the default replica number is not 1 and whether the HDFS service is normal.

HBase High-Risk Operations

Table 12-8 HBase high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify encryption configuration. ● hbase.regionserver.wal.encryption ● hbase.crypto.keyprovider.parameters.uri ● hbase.crypto.keyprovider.parameters.encryptedtext	Services cannot start properly.	▲▲▲▲	Strictly follow the prompt information when modifying related configuration items, which are associated. Ensure that new values are valid.	Check whether services can be started properly.

Operation	Risk	Severity	Workaround	Check Item
Change the value of **hbase.regionserver.wal.encryption** to **false** or switch encryption algorithm from AES to SMS4.	This operation may cause start failures and data loss.	▲ ▲ ▲ ▲	When HFile and WAL are encrypted using an encryption algorithm and a table is created, do not close or switch the encryption algorithm randomly. If an encryption table (ENCRYPTION=>AES/SMS4) is not created, you can only switch the encryption algorithm.	None
Modify HBase instance start parameter **GC_OPTS** and **HBASE_HEAPSIZE**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid. GC_OPTS does not conflict with HBASE_HEAPSIZE.	Check whether services can be started properly.
Use **OfflineMetaRepair** tool	Services cannot start properly.	▲ ▲ ▲ ▲	This tool can be used only when HBase is offline and cannot be used in data migration scenarios.	Check whether HBase services can be started properly.

HDFS High-Risk Operations

Table 12-9 HDFS high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Change HDFS NameNode data storage directory **dfs.namenode.name.dir** and data configuration directory **dfs.datanode.data.dir**.	Services cannot start properly.	▲▲▲▲▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Use the **-delete** parameter when you run the **hadoop distcp** command.	During DistCP copying, files that do not exist in the source cluster but exist in the destination cluster are deleted from the destination cluster.	▲▲	When using DistCP, determine whether to retain the redundant files in the destination cluster. Exercise caution when using the **-delete** parameter.	After DistCP copying is complete, check whether the data in the destination cluster is retained or deleted according to the parameter settings.

Operation	Risk	Severity	Workaround	Check Item
Modify the HDFS instance start parameter **GC_OPTS**, **HADOOP_HEAPSIZE**, and **GC_PROFILE**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid. GC_OPTS does not conflict with HADOOP_HEAPSIZE.	Check whether services can be started properly.
Change the default value of **dfs.replication** from **3** to **1**.	This operation will have the following impacts: 1. The storage reliability deteriorates. If the disk becomes faulty, data will be lost. 2. NameNode fails to be restarted, and the HDFS service is unavailable.	▲ ▲ ▲ ▲	When modifying related configuration items, check the parameter description carefully. Ensure that there are more than two replicas for data storage.	Check whether the default replica number is not 1 and whether the HDFS service is normal.
Change the remote procedure call (RPC) channel encryption mode (**hadoop.rpc.protection**) of each module in Hadoop.	This operation causes service faults and service exceptions.	▲ ▲ ▲ ▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether HDFS and other services that depend on HDFS can properly start and provide services.

Hive High-Risk Operations

Table 12-10 Hive high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify the Hive instance start parameter **GC_OPTS**.	This operation may cause Hive instance start failures.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Delete all MetaStore instances.	This operation may cause Hive metadata loss. As a result, Hive cannot provide services.	▲ ▲ ▲	Do not perform this operation unless ensure that Hive table information can be discarded.	Check whether services can be started properly.
Delete or modify files corresponding to Hive tables over HDFS interfaces or HBase interfaces.	This operation may cause Hive service data loss or tampering.	▲ ▲	Do not perform this operation unless ensure that the data can be discarded or that the operation meets service requirements.	Check whether Hive data is complete.

Operation	Risk	Severity	Workaround	Check Item
Delete or modify files corresponding to Hive tables or directory access permission over HDFS interfaces or HBase interfaces.	This operation may cause related service scenarios to be unavailable.	▲ ▲ ▲	Do not perform this operation.	Check whether related service operations are normal.
Delete or modify **hdfs:///apps/templeton/hive-3.1.0.tar.gz** over HDFS interfaces.	WebHCat fails to perform services due to this operation.	▲ ▲	Do not perform this operation.	Check whether related service operations are normal.
Export table data to overwrite the data at the local. For example, export the data of **t1** to **/opt/dir**. **insert overwrite local directory '/opt/dir' select * from t1;**	This operation will delete target directories. Incorrect setting may cause software or OS startup failures.	▲ ▲ ▲ ▲ ▲	Ensure that the path where the data is written does not contain any files or do not use the key word overwrite in the command.	Check whether files in the target path are lost.

Operation	Risk	Severity	Workaround	Check Item
Direct different databases, tables, or partition files to the same path, for example, default warehouse path **/user/hive/warehouse**.	The creation operation may cause disordered data. After a database, table, or partition is deleted, other object data will be lost.	▲ ▲ ▲ ▲ ▲	Do not perform this operation.	Check whether files in the target path are lost.

Kafka High-Risk Operations

Table 12-11 Kafka high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Delete Topic	This operation may delete existing topics and data.	▲ ▲ ▲	Kerberos authentication is used to ensure that authenticated users have operation permissions. Ensure that topic names are correct.	Check whether topics are processed properly.
Delete data directories.	This operation may cause service information loss.	▲ ▲ ▲	Do not delete data directories manually.	Check whether data directories are normal.

Operation	Risk	Severity	Workaround	Check Item
Modify data directory content (file and folder creation).	This operation may cause the Broker instance of the node faults.	▲ ▲ ▲	Do not create or modify files or folders in the data directories manually.	Check whether data directories are normal.
Modify the disk auto-adaptation function using the **disk.adapter.enable** parameter.	This operation adjusts the topic data retention period when the disk usage reaches the threshold. Historical data that does not fall within the storage retention may be deleted.	▲ ▲ ▲	If the retention period of some topics cannot be adjusted, add this topic to the value of **disk.adapter.topic.blacklist**.	Observe the data storage period on the Kafka topic monitoring page.
Modify data directory **log.dirs** configuration.	Incorrect operation may cause process faults.	▲ ▲ ▲	Ensure that the added or modified data directories are empty and that the directory permissions are right.	Check whether data directories are normal.
Reduce the capacity of the Kafka cluster.	This operation may cause quantity reduction of backups of some data duplicates of topic. As a result, some topics cannot be accessed.	▲ ▲	Perform backup operation and then reduce the capacity of the Kafka cluster.	Check whether backup nodes where partitions are located are activated to ensure data security.

Operation	Risk	Severity	Workaround	Check Item
Start or stop basic components independently.	This operation has adverse impact on the basic functions of some services. As a result, service failures occur.	▲ ▲ ▲	Do not start or stop ZooKeeper, Kerberos, and LDAP basic components independently. Select related services when performing this operation.	Check whether the service status is normal.
Restart or stop services.	This operation may interrupt services.	▲ ▲	Restart or stop services when necessary.	Check whether the service is running properly.
Modify configuration parameters.	This operation requires service restart for configuration to take effect.	▲ ▲	Modify configuration when necessary.	Check whether the service is running properly.
Delete or modify metadata.	Modifying or deleting Kafka metadata on ZooKeeper may cause the Kafka topic or service unavailability.	▲ ▲ ▲	Do not delete or modify Kafka metadata stored on ZooKeeper.	Check whether the Kafka topics or Kafka service is available.
Delete metadata backup files.	After Kafka metadata backup files are modified and used to restore Kafka metadata, Kafka topics or the Kafka service may be unavailable.	▲ ▲ ▲	Do not delete Kafka metadata backup files.	Check whether the Kafka topics or Kafka service is available.

KrbServer High-Risk Operations

Table 12-12 KrbServer high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify the **KADMIN_PORT** parameter of KrbServer.	After this parameter is modified, if the KrbServer service and its associated services are not restarted in a timely manner, the configuration of KrbClient in the cluster is abnormal and the service running is affected.	▲▲▲▲▲	After this parameter is modified, restart the KrbServer service and all its associated services.	None
Modify the **kdc_ports** parameter of KrbServer.	After this parameter is modified, if the KrbServer service and its associated services are not restarted in a timely manner, the configuration of KrbClient in the cluster is abnormal and the service running is affected.	▲▲▲▲▲	After this parameter is modified, restart the KrbServer service and all its associated services.	None
Modify the **KPASSWD_PORT** parameter of KrbServer.	After this parameter is modified, if the KrbServer service and its associated services are not restarted in a timely manner, the configuration of KrbClient in the cluster is abnormal and the service running is affected.	▲▲▲▲▲	After this parameter is modified, restart the KrbServer service and all its associated services.	None

Operation	Risk	Severity	Workaround	Check Item
Modify the domain name of Manager system.	After the domain name is modified, if the KrbServer service and its associated services are not restarted in a timely manner, the configuration of KrbClient in the cluster is abnormal and the service running is affected.	▲▲▲▲▲	After this parameter is modified, restart the KrbServer service and all its associated services.	None
Configure cross-cluster mutual trust relationships.	This operation will restart the KrbServer service and all associated services, affecting the management and maintenance and services of the cluster.	▲▲▲▲▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.

LdapServer High-Risk Operations

Table 12-13 LdapServer high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify the **LDAP_SERVER_PORT** parameter of LdapServer.	After this parameter is modified, if the LdapServer service and its associated services are not restarted in a timely manner, the configuration of LdapClient in the cluster is abnormal and the service running is affected.	▲▲▲▲▲	After this parameter is modified, restart the LdapServer service and all its associated services.	None

Operation	Risk	Severity	Workaround	Check Item
Restore LdapServer data.	This operation will restart Manager and the entire cluster, affecting management, maintenance, and services of the cluster.	▲ ▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Replace the Node where LdapServer is located.	This operation will interrupt services deployed on the node. If the node is a management node, the operation will restart all OMS processes, affecting the cluster management and maintenance.	▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Change the password of LdapServer.	The LdapServer and Kerberos services need to be restarted during the password change, affecting the management, maintenance, and services of the cluster.	▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	None

Operation	Risk	Severity	Workaround	Check Item
Restart the node where LdapServer is located.	Restarting the node without stopping the LdapServer service may cause LdapServer data damage.	▲▲▲▲▲	Restore LdapServer using LdapServer backup data	None

Loader High-Risk Operations

Table 12-14 Loader high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Change the floating IP address of a Loader instance (**loader.float.ip**).	Services cannot start properly.	▲▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether the Loader UI can be connected properly.
Modify the Loader instance start parameter **LOADER_GC_OPTS**.	Services cannot start properly.	▲▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Clear table contents when adding data to HBase.	This operation will clear original data in the target table.	▲▲	Ensure that the contents in the target table can be cleared before the operation.	Check whether the contents in the target table can be cleared before the operation.

Spark2x High-risk Operations

NOTE

Spark high-risk operations apply to MRS 3.x earlier versions.

Table 12-15 Spark2x high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify the configuration item **spark.yarn.queue**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Modify the configuration item **spark.driver.extraJavaOptions**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Modify the configuration item **spark.yarn.cluster.driver.extraJavaOptions**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.

Operation	Risk	Severity	Workaround	Check Item
Modify the configuration item **spark.eventLog.dir**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Modify the configuration item **SPARK_DAEMON_JAVA_OPTS**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Delete all JobHistory2x instances.	The event logs of historical applications are lost.	▲ ▲	Reserve at least one JobHistory2x instance.	Check whether historical application information is included in JobHistory2x.
Delete or modify the **/user/spark2x/jars/8.1.0.1/spark-archive-2x.zip** file in HDFS.	JDBCServer2x fails to be started and service functions are abnormal.	▲ ▲ ▲	Delete **/user/spark2x/jars/8.1.0.1/spark-archive-2x.zip**, and wait for 10-15 minutes until the .zip package is automatically restored.	Check whether services can be started properly.

Storm High-Risk Operations

Table 12-16 Storm high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify the following plug-in related configuration items: ● storm.scheduler ● nimbus.authorizer ● storm.thrift.transport ● nimbus.blobstore.class ● nimbus.topology.validator ● storm.principal.tolocal	Services cannot start properly.	▲▲▲▲	Strictly follow the prompt information when modifying related configuration items. Ensure that the class names exist and are valid.	Check whether services can be started properly.

Operation	Risk	Severity	Workaround	Check Item
Modify the Storm instance **GC_OPTS** startup parameters, including: NIMBUS_GC_OPTS SUPERVISOR_GC_OPTS UI_GC_OPTS LOGVIEWER_GC_OPTS	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Modify the user resource pool configuration parameter **resource.aware.scheduler.user.pools**.	Services cannot run properly.	▲ ▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that resources allocated to each user are appropriate and valid.	Check whether services can be started and run properly
Change data directories.	If this operation is not properly performed, services may be abnormal and unavailable.	▲ ▲ ▲ ▲	Do not manually change data directories.	Check whether data directories are normal.
Restart services or instances.	The service will be interrupted for a short period of time, and ongoing operations will be interrupted.	▲ ▲ ▲	Restart services or instances when necessary.	Check whether the service is running properly and whether interrupted operations are restored.

Operation	Risk	Severity	Workaround	Check Item
Synchronize configurations (by restarting the required service).	The service will be restarted, resulting in temporary service interruption. If Supervisor is restarted, ongoing operations will be interrupted for a short period of time.	▲ ▲ ▲	Modify configuration when necessary.	Check whether the service is running properly and whether interrupted operations are restored.
Stop services or instances.	The service will be stopped, and related operations will be interrupted.	▲ ▲ ▲	Stop services when necessary.	Check whether the services are properly stopped.
Delete or modify metadata.	If Nimbus metadata is deleted, services are abnormal and ongoing operations are lost.	▲ ▲ ▲ ▲ ▲	Do not manually delete Nimbus metadata files.	Check whether Nimbus metadata files are normal.
Modify file permissions.	If permissions on the metadata and log directories are incorrectly modified, service exceptions may occur.	▲ ▲ ▲ ▲	Do not manually modify file permissions.	Check whether the permissions on the data and log directories are correct.
Delete topologies.	Topologies in use will be deleted.	▲ ▲ ▲ ▲	Delete topologies when necessary.	Check whether the topologies are successfully deleted.

Yarn High-Risk Operations

Table 12-17 Yarn high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Delete or change data directories **yarn.nodemanager.local-dirs** and **yarn.nodemanager.log-dirs**	This operation may cause service information loss.	▲▲▲	Do not delete data directories manually.	Check whether data directories are normal.

ZooKeeper High-Risk Operations

Table 12-18 ZooKeeper high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Delete or change ZooKeeper data directories.	This operation may cause service information loss.	▲▲▲	Follow the capacity expansion guide to change the ZooKeeper data directories.	Check whether services and associated components are started properly.
Modify the ZooKeeper instance start parameter **GC_OPTS**.	Services cannot start properly.	▲▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.

Operation	Risk	Severity	Workaround	Check Item
Modify the znode ACL information in ZooKeeper.	If znode permission is modified in ZooKeeper, other users may have no permission to access the znode and some system functions are abnormal.	▲ ▲ ▲ ▲	During the modification, strictly follow the ZooKeeper Configuration Guide and ensure that other components can use ZooKeeper properly after ACL information modification.	Check that other components that depend on ZooKeeper can properly start and provide services.

13 MRS Quick Start

13.1 How to Use MRS

MapReduce Service (MRS) is used to deploy and manage the Hadoop system and enables one-click Hadoop cluster deployment. MRS provides enterprise-level big data clusters on the cloud. Tenants can fully control the clusters and easily run big data components such as Hadoop, Spark, HBase, Kafka, and Storm in the clusters.

MRS is easy to use. You can execute various tasks and process or store PB-level data using computers connected in a cluster. The procedure of using MRS is as follows:

1. Upload local programs and data files to OBS.
2. Create a cluster by following instructions in **Creating a Custom Cluster**. You can choose a cluster type for offline data analysis or stream processing or both, and set ECS instance specifications, instance count, data disk type (common I/O, high I/O, and ultra-high I/O), and components to be installed such as Hadoop, Spark, HBase, Hive, Kafka, and Storm in a cluster. You can use a **bootstrap action** to execute a script on a specified node before or after the cluster is started to install additional third-party software, modify the cluster running environment, and perform other customizations.
3. **Manage jobs**. MRS provides a platform for executing programs you develop. You can submit, execute, and monitor such programs on MRS.
4. **Manage clusters**. MRS provides you with MRS Manager, an enterprise-level unified management platform of big data clusters, helping you quickly know health status of services and hosts. Through graphical metric monitoring and customization, you can obtain critical system information in a timely manner. In addition, you can modify service attribute configurations based on service performance requirements, and start or stop clusters, services, and role instances in one click.
5. **Terminate a cluster**. You can terminate an MRS cluster that is no longer use after job execution is complete.

13.2 Creating a Cluster

The first step of using MRS is to create a cluster. This section describes how to create a cluster on the MRS management console.

Procedure

Step 1 Log in to the MRS console.

Step 2

NOTE

When creating a cluster, pay attention to quota notification. If a resource quota is insufficient, increase the resource quota as prompted and create a cluster.

Step 3 On the page for create a cluster, click the **Custom Config** tab.

Step 4 Configure cluster software information.

- **Region**: Use the default value.
- **Cluster Name**: You can use the default name. However, you are advised to include a project name abbreviation or date for consolidated memory and easy distinguishing, for example, **mrs_20180321**.
- **Cluster Version**: Select the latest version, which is the default value.
- **Cluster Type**: Use the default **Analysis Cluster**.
- **Component**: Select components such as Spark2x, HBase, and Hive for the analysis cluster. For a streaming cluster, select components such as Kafka and Storm. For a hybrid cluster, you can select the components of the analysis cluster and streaming cluster based on service requirements.
- **Metadata**: Retain the default value.

  NOTE

  For versions earlier than MRS 3.x, select components such as Spark, HBase, and Hive for the analysis cluster.

Step 5 Click **Next**.

- **AZ**: Use the default value.
- **VPC**: Use the default value. If there is no available VPC, click **View VPC** to access the VPC console and create a new VPC.
- **Subnet**: Use the default value.
- **Security Group**: Select **Auto create**.
- **EIP**: Select **Bind later**.
- **Instance Specifications**: Select General Computing S3 -> 8 vCPUs | 16 GB (s3.2xlarge.2) for both Master and Core nodes.
- **System Disk**: Select **Common I/O** and retain the default settings.
- **Data Disk**: Select **Common I/O** and retain the default settings.
- **Instance Count**: The default number of Master nodes is 2, and that of Core nodes is 3.

Step 6 Click **Next**. The **Set Advanced Options** tab page is displayed. Configure the following parameters. Retain the default settings for the other parameters.

- Kerberos authentication:
  - **Kerberos Authentication**: Disable Kerberos authentication.
  - **Username**: name of the Manager administrator. **admin** is used by default.
  - **Password**: password of the Manager administrator.
- **Key Pair**: Select a key pair from the drop-down list. Select **"I acknowledge that I have obtained private key file** ***SSHkey-xxx*** **and that without this file I will not be able to log in to my ECS."** If you have never created a key pair, click **View Key Pair** to create or import a key pair. And then, obtain a private key file.
- **Secure Communications**: Select **Enable**.

Step 7 Click **Apply Now**.

If Kerberos authentication is enabled for a cluster, check whether Kerberos authentication is required. If yes, click **Continue**. If no, click **Back** to disable Kerberos authentication and then create a cluster.

Step 8 Click **Back to Cluster List** to view the cluster status.

It takes some time to create a cluster. The initial status of the cluster is **Starting**. After the cluster has been created successfully, the cluster status becomes **Running**.

----End

13.3 Uploading Data and Programs

Through the **Files** tab page, you can create, delete, import, export, delete files in the analysis cluster.

Background

MRS clusters process data from OBS or HDFS. OBS provides customers with the data storage capabilities that are massive, secure, reliable, and cost-effective. MRS can directly process data in OBS. You can browse, manage, and use data on the web page of the management console and OBS Client.

Importing Data

Currently, MRS can only import data from OBS to HDFS. The file upload rate decreases with the increase of the file size. This mode applies to scenarios where the data volume is small.

You can perform the following steps to import files and directories:

1. Log in to the MRS console.
2. Choose **Clusters > Active Clusters** and click the name of the cluster to be queried to enter the page displaying the cluster's information.
3. Click the **Files** tab to go to the file management page.

4. Select **HDFS File List**.
5. Go to the data storage directory, for example, **bd_app1**.

   The **bd_app1** directory is only an example. You can use any directory on the page or create a new one.

   The requirements for creating a folder are as follows:

   – The folder name contains a maximum of 255 characters
   – The folder name cannot be empty.
   – The folder name cannot contain the following special characters: /:*?"<>|\;&,'`!{}[]$%+
   – The value cannot start or end with a period (.).
   – The spaces at the beginning and end are ignored.
6. Click **Import Data** and configure the HDFS and OBS paths correctly. When configuring the OBS or HDFS path, click **Browse**, select a file directory, and click **Yes**.
   – OBS path
     ■ The path must start with **obs://**.
     ■ Files or programs encrypted by KMS cannot be imported.
     ■ An empty folder cannot be imported.
     ■ The directory and file name can contain letters, digits, hyphens (-), and underscores (_), but cannot contain the following special characters ;|&>,<'$*?\
     ■ The directory and file name cannot start or end with a space, but can contain spaces between them.
     ■ The OBS full path contains a maximum of characters.
   – HDFS path
     ■ The path starts with **/user** by default.
     ■ The directory and file name can contain letters, digits, hyphens (-), and underscores (_), but cannot contain the following special characters: ;|&>,<'$*?\:
     ■ The directory and file name cannot start or end with a space, but can contain spaces between them.
     ■ The HDFS full path contains a maximum of characters.
7. Click **OK**.

   You can view the file upload progress on the **File Operation Records** tab page. MRS processes the data import operation as a DistCp job. You can also check whether the DistCp job is successfully executed on the **Jobs** tab page.

Exporting Data

After the data analysis and computing are completed, you can store the data in HDFS or export them to OBS.

You can perform the following steps to export files and directories:

1. Log in to the MRS console.
2. Choose **Clusters > Active Clusters** and click the name of the cluster to be queried to enter the page displaying the cluster's basic information.
3. Click the **Files** tab to go to the file management page.
4. Select **HDFS File List**.
5. Go to the data storage directory, for example, **bd_app1**.
6. Click **Export Data** and configure the OBS and HDFS paths. When configuring the OBS or HDFS path, click **Browse**, select a file directory, and click **Yes**.
   - OBS path
     - The path must start with **obs://**.
     - The directory and file name can contain letters, digits, hyphens (-), and underscores (_), but cannot contain the following special characters ;|&>,<'$*?\
     - The directory and file name cannot start or end with a space, but can contain spaces between them.
     - The OBS full path contains a maximum of characters.
   - HDFS path
     - The path starts with **/user** by default.
     - The directory and file name can contain letters, digits, hyphens (-), and underscores (_), but cannot contain the following special characters: ;|&>,<'$*?\:
     - The directory and file name cannot start or end with a space, but can contain spaces between them.
     - The HDFS full path contains a maximum of characters.

   NOTE

   When a folder is exported to OBS, a label file named **folder name_$folder$** is added to the OBS path. Ensure that the exported folder is not empty. If the exported folder is empty, OBS cannot display the folder and only generates a file named **folder name_ $folder$**.
7. Click **OK**.

   You can view the file upload progress on the **File Operation Records** tab page. MRS processes the data export operation as a DistCp job. You can also check whether the DistCp job is successfully executed on the **Jobs** tab page.

13.4 Creating a Job

You can submit programs developed by yourself to MRS to execute them, and obtain the results.

This section describes how to submit a job (take a MapReduce job as an example) on the MRS management console. MapReduce jobs are used to submit JAR programs to quickly process massive amounts of data in parallel and create a distributed data processing and execution environment.

If the job and file management functions are not supported on the cluster details page, submit the jobs in the background.

Before creating a job, you need to upload local data to OBS for data computing and analyzing. MRS allows exporting data from OBS to HDFS for computing and analyzing. After the analyzing and computing are complete, you can store the data in HDFS or export them to OBS. HDFS and OBS can also store the compressed data in the format of **bz2** or **gz**.

Submitting a Job on the GUI

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters > Active Clusters**, select a running cluster, and click its name to switch to the cluster details page.

Step 3 If Kerberos authentication is enabled for the cluster, perform the following steps. If Kerberos authentication is not enabled for the cluster, skip this step.

In the **Basic Information** area on the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users. For details, see **Synchronizing IAM Users to MRS**.

NOTE

- When the policy of the user group to which the IAM user belongs changes from MRS ReadOnlyAccess to MRS CommonOperations, MRS FullAccess, or MRS Administrator, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** (System Security Services Daemon) cache of cluster nodes needs time to be updated. Then, submit a job. Otherwise, the job may fail to be submitted.
- When the policy of the user group to which the IAM user belongs changes from MRS CommonOperations, MRS FullAccess, or MRS Administrator to MRS ReadOnlyAccess, wait for 5 minutes until the new policy takes effect after the synchronization is complete because the **SSSD** cache of cluster nodes needs time to be updated.

Step 4 Click the **Jobs** tab.

Step 5 Click **Create**. The **Create Job** page is displayed.

NOTE

If the IAM username contains spaces (for example, **admin 01**), a job cannot be created.

Step 6 In **Type**, select **MapReduce**. Configure other job information.

- Configure MapReduce job information by referring to **Table 13-3** if the cluster version is earlier than MRS 2.1.0.

Table 13-1 Job configuration information

Parameter	Description
Name	Job name. It contains 1 to 64 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed. **NOTE** You are advised to set different names for different jobs.
Program Path	Path of the program package to be executed. The following requirements must be met: ● Contains a maximum of 1,023 characters, excluding special characters such as ;|&><'$. The parameter value cannot be empty or full of spaces. ● The path of the program to be executed can be stored in HDFS or OBS. The path varies depending on the file system. – OBS: The path must start with **obs://**. Example: **obs://wordcount/program/*xxx*.jar** – HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. ● For SparkScript and HiveScript, the path must end with **.sql**. For MapReduce, the path must end with **.jar**. For Flink and SparkSubmit, the path must end with **.jar** or **.py**. The **.sql**, **.jar**, and **.py** are case-insensitive.
Parameters	(Optional) It is the key parameter for program execution. Multiple parameters are separated by space. Configuration method: *Program class name Data input path Data output path* ● Program class name: It is specified by a function in your program. MRS is responsible for transferring parameters only. ● Data input path: Click **HDFS** or **OBS** to select a path or manually enter a correct path. ● Data output path: Enter a directory that does not exist. The parameter contains a maximum of 150,000 characters. It cannot contain special characters ;|&><'$, but can be left blank. **CAUTION** If you enter a parameter with sensitive information (such as the login password), the parameter may be exposed in the job details display and log printing. Exercise caution when performing this operation.
Service Parameter	(Optional) It is used to modify service parameters for the job. The parameter modification applies only to the current job. To make the modification take effect permanently for the cluster, follow instructions in **Configuring Service Parameters**. To add multiple parameters, click ⊕ on the right. To delete a parameter, click **Delete** on the right. **Table 13-2** lists the common service configuration parameters.

Parameter	Description
Command Reference	Command submitted to the background for execution when a job is submitted.

Table 13-2 Service Parameter parameters

Parameter	Description	Example Value
fs.obs.access.key	Key ID for accessing OBS.	-
fs.obs.secret.key	Key corresponding to the key ID for accessing OBS.	-

Table 13-3 Job configuration information

Parameter	Description
Name	Job name. It contains 1 to 64 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed. **NOTE** You are advised to set different names for different jobs.
Program Path	Path of the program package to be executed. The following requirements must be met: ● Contains a maximum of 1,023 characters, excluding special characters such as ;|&><'$. The parameter value cannot be empty or full of spaces. ● The path of the program to be executed can be stored in HDFS or OBS. The path varies depending on the file system. – OBS: The path must start with **s3a://**. Example: **s3a://wordcount/program/xxx.jar** – HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. ● For SparkScript, the path must end with **.sql**. For MapReduce and Spark, the path must end with **.jar**. The **.sql** and **.jar** are case-insensitive.

Parameter	Description
Parameters	Key parameter for program execution. The parameter is specified by the function of the user's program. MRS is only responsible for loading the parameter. Multiple parameters are separated by space. Configuration method: *Package name*.*Class name* The parameter contains a maximum of 150,000 characters. It cannot contain special characters ;|&><'$, but can be left blank. **NOTE** When entering a parameter containing sensitive information (for example, login password), you can add an at sign (@) before the parameter name to encrypt the parameter value. This prevents the sensitive information from being persisted in plaintext. When you view job information on the MRS management console, the sensitive information is displayed as *****. Example: **username=admin @password=admin_123**
Import From	Path for inputting data Data can be stored in HDFS or OBS. The path varies depending on the file system. ● OBS: The path must start with **s3a://**. ● HDFS: The path must start with **/user**. For details about how to import data to HDFS, see **Importing Data**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.
Export To	Path for outputting data **NOTE** ● When setting this parameter, select **OBS** or **HDFS**. Select a file directory or manually enter a file directory, and click **OK**. ● If you add the **hadoop-mapreduce-examples-x.x.x.jar** sample program or a program similar to **hadoop-mapreduce-examples-x.x.x.jar**, enter a directory that does not exist. Data can be stored in HDFS or OBS. The path varies depending on the file system. ● OBS: The path must start with **s3a://**. (Supported only in MRS 1.8.10 and earlier versions) ● HDFS: The path must start with **/user**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.

Parameter	Description
Log Path	Path for storing job logs that record job running status. Data can be stored in HDFS or OBS. The path varies depending on the file system. ● OBS: The path must start with **s3a://**. ● HDFS: The path must start with **/user**. The parameter contains a maximum of 1,023 characters, excluding special characters such as ;|&>,<'$, and can be left blank.

Step 7 Confirm job configuration information and click **OK**.

After the job is created, you can manage it.

----End

Submitting a Job in the Background

The default client installation path for MRS 3.x or later is /opt/Bigdata/client, and for versions earlier than MRS 3.x is /opt/client. Configure the path based on site requirements.

Step 1 Log in to a Master node. For details, see **Logging In to an ECS**.

Step 2 Run the following command to initialize environment variables:

source /opt/Bigdata/client/bigdata_env

Step 3 If the Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If the Kerberos authentication is disabled for the current cluster, skip this step.

kinit *MRS cluster user*

Example: **kinit admin**

Step 4 Run the following command to copy the program in the OBS file system to the Master node in the cluster:

hadoop fs -Dfs.obs.access.key=AK -Dfs.obs.secret.key=SK -copyToLocal source_path.jar target_path.jar

Example: **hadoop fs -Dfs.obs.access.key=XXXX -Dfs.obs.secret.key=XXXX -copyToLocal "obs://mrs-word/program/hadoop-mapreduce-examples-XXX.jar" "/home/omm/hadoop-mapreduce-examples-XXX.jar"**

You can log in to OBS Console using AK/SK. To obtain AK/SK information, click the username in the upper right corner of the management console and choose **My Credentials** > **Access Keys**.

Step 5 Run the following command to submit a wordcount job. If data needs to be read from OBS or outputted to OBS, the AK/SK parameters need to be added.

source /opt/Bigdata/client/bigdata_env;hadoop jar execute_jar wordcount input_path output_path

Example: **source /opt/Bigdata/client/bigdata_env;hadoop jar /home/omm/hadoop-mapreduce-examples-XXX.jar wordcount -Dfs.obs.access.key=XXXX -Dfs.obs.secret.key=XXXX "obs://mrs-word/input/*" "obs://mrs-word/output/"**

In the preceding command, **input_path** indicates a path for storing job input files on OBS. **output_path** indicates a path for storing job output files on OBS and needs to be set to a directory that does not exist

----End

13.5 Using Clusters with Kerberos Authentication Enabled

This section instructs you to use security clusters and run MapReduce, Spark, and Hive programs.

The Presto component of MRS 3.x does not support Kerberos authentication.

You can get started by reading the following topics:

1. **Creating a Security Cluster and Logging In to Manager**
2. **Creating a Role and a User**
3. **Running a MapReduce Program**
4. **Running a Spark Program**
5. **Running a Hive Program**

Creating a Security Cluster and Logging In to Manager

Step 1 Create a security cluster. For details, see **Creating a Custom Cluster**. Enable **Kerberos Authentication**, set **Password**, and confirm the password. This password is used to log in to Manager. Keep it secure.

Step 2 Log in to the MRS console.

Step 3 In the navigation pane on the left, choose **Active Clusters** and click the target cluster name on the right to access the cluster details page.

Step 4 Click **Access Manager** on the right of **MRS Manager** to log in to Manager.

- If you have bound an EIP when creating the cluster, perform the following operations:

  a. Add a security group rule. By default, your public IP address used for accessing port 9022 is filled in the rule. If you want to view, modify, or delete a security group rule, click **Manage Security Group Rule**.

     NOTE

     - It is normal that the automatically generated public IP address is different from your local IP address and no action is required.
     - If port 9022 is a Knox port, you need to enable the permission to access port 9022 of Knox for accessing Manager.

  b. Select **I confirm that xx.xx.xx.xx is a trusted public IP address and MRS Manager can be accessed using this IP address.**

- If you have not bound an EIP when creating the cluster, perform the following operations:

    a. Select an available EIP from the drop-down list or click **Manage EIP** to create one.

    b. Add a security group rule. By default, your public IP address used for accessing port 9022 is filled in the rule. If you want to view, modify, or delete a security group rule, click **Manage Security Group Rule**.

    NOTE

    - It is normal that the automatically generated public IP address is different from the local IP address and no action is required.
    - If port 9022 is a Knox port, you need to enable the permission of port 9022 to access Knox for accessing MRS Manager.

    c. Select **I confirm that xx.xx.xx.xx is a trusted public IP address and MRS Manager can be accessed using this IP address.**

Step 5 Click **OK**. The Manager login page is displayed. To assign permissions to other users to access Manager, add their public IP addresses as trusted ones by referring to **Accessing MRS Manager MRS 2.x or Earlier)**.

Step 6 Enter the default username **admin** and the password you set when creating the cluster, and click **Log In**.

----End

Creating a Role and a User

For clusters with Kerberos authentication enabled, perform the following steps to create a user and assign permissions to the user to run programs.

Step 1 On Manager, choose **System** > **Permission** > **Role**.

Step 2 Click **Create Role**. For details, see **Creating a Role**.

Specify the following information:

- Enter a role name, for example, **mrrole**.
- In **Configure Resource Permission**, select the cluster to be operated, choose **Yarn** > **Scheduler Queue** > **root**, and select **Submit** and **Admin** in the **Permission** column. After you finish configuration, do not click **OK** but click the name of the target cluster shown in the following figure and then configure other permissions.
- Choose **HBase** > **HBase Scope**. Locate the row that contains **global**, and select **create**, **read**, **write**, and **execute** in the **Permission** column. After you finish configuration, do not click **OK** but click the name of the target cluster shown in the following figure and then configure other permissions.
- Choose **HDFS** > **File System** > **hdfs://hacluster/** and select **Read**, **Write**, and **Execute** in the **Permission** column. After you finish configuration, do not click **OK** but click the name of the target cluster shown in the following figure and then configure other permissions.
- Choose **Hive** > **Hive Read Write Privileges**, select **Select**, **Delete**, **Insert**, and **Create** in the **Permission** column, and click **OK**.

Step 3 Choose **System**. In the navigation pane on the left, choose **Permission** > **User Group** > **Create User Group** to create a user group for the sample project, for example, **mrgroup**. For details, see **Creating a User Group**.

Step 4 Choose **System**. In the navigation pane on the left, choose **Permission** > **User** > **Create** to create a user for the sample project. For details, see **Creating a User**.

- Enter a username, for example, **test**. If you want to run a Hive program, enter **hiveuser** in **Username**.
- Set **User Type** to **Human-Machine**.
- Enter a password. This password will be used when you run the program.
- In **User Group**, add **mrgroup** and **supergroup**.
- Set **Primary Group** to **supergroup** and bind the **mrrole** role to obtain the permission.

  Click **OK**.

Step 5 Choose **System**. In the navigation pane on the left, choose **Permission** > **User**, locate the row where user **test** locates, and select **Download Authentication Credential** from the **More** drop-down list. Save the downloaded package and decompress it to obtain the **keytab** and **krb5.conf** files.

----End

Running a MapReduce Program

This section describes how to run a MapReduce program in security cluster mode.

Prerequisites

You have compiled the program and prepared data files, for example, **mapreduce-examples-1.0.jar**, **input_data1.txt**, and **input_data2.txt**..

Procedure

Step 1 Use a remote login software (for example, MobaXterm) to log in to the master node of the security cluster using SSH (using the EIP).

Step 2 After the login is successful, run the following commands to create the **test** folder in the **/opt/Bigdata/client** directory and create the **conf** folder in the **test** directory:

```
cd /opt/Bigdata/client
mkdir test
cd test
mkdir conf
```

Step 3 Use an upload tool (for example, WinSCP) to copy **mapreduce-examples-1.0.jar**, **input_data1.txt**, and **input_data2.txt** to the **test** directory, and copy the **keytab** and **krb5.conf** files obtained in **Step 5** in **Creating Roles and Users** to the **conf** directory.

Step 4 Run the following commands to configure environment variables and authenticate the created user, for example, **test**:

```
cd /opt/Bigdata/client
source bigdata_env
export YARN_USER_CLASSPATH=/opt/Bigdata/client/test/conf/
kinit test
```

Enter the password as prompted. If no error message is displayed (you need to change the password as prompted upon the first login), Kerberos authentication is complete.

Step 5 Run the following commands to import data to the HDFS:

```
cd test
hdfs dfs -mkdir /tmp/input
hdfs dfs -put input_data* /tmp/input
```

Step 6 Run the following commands to run the program:

```
yarn jar mapreduce-examples-1.0.jar com.huawei.bigdata.mapreduce.examples.FemaleInfoCollector /tmp/
input /tmp/mapreduce_output
```

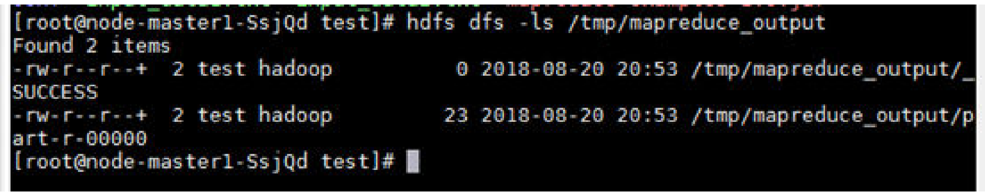
In the preceding commands:

/tmp/input indicates the input path in the HDFS.

/tmp/mapreduce_output indicates the output path in the HDFS. This directory must not exist. Otherwise, an error will be reported.

Step 7 After the program is executed successfully, run the **hdfs dfs -ls /tmp/mapreduce_output** command. The following command output is displayed.

Figure 13-1 Program running result



```
[root@node-master1-SsjQd test]# hdfs dfs -ls /tmp/mapreduce_output
Found 2 items
-rw-r--r--+  2 test hadoop          0 2018-08-20 20:53 /tmp/mapreduce_output/_
SUCCESS
-rw-r--r--+  2 test hadoop         23 2018-08-20 20:53 /tmp/mapreduce_output/p
art-r-00000
[root@node-master1-SsjQd test]#
```

----End

Running a Spark Program

This section describes how to run a Spark program in security cluster mode.

Prerequisites

You have compiled the program and prepared data files, for example, **FemaleInfoCollection.jar**, **input_data1.txt**, and **input_data2.txt**..

Procedure

Step 1 Use a remote login software (for example, MobaXterm) to log in to the master node of the security cluster using SSH (using the EIP).

Step 2 After the login is successful, run the following commands to create the **test** folder in the **/opt/Bigdata/client** directory and create the **conf** folder in the **test** directory:

```
cd /opt/Bigdata/client
mkdir test
cd test
mkdir conf
```

Step 3 Use an upload tool (for example, WinSCP) to copy **FemaleInfoCollection.jar**, **input_data1.txt**, and **input_data2.txt** to the **test** directory, and copy the **keytab** and **krb5.conf** files obtained in **Step 5** in section **Creating Roles and Users** to the **conf** directory.

Step 4 Run the following commands to configure environment variables and authenticate the created user, for example, **test**:

```
cd /opt/Bigdata/client
source bigdata_env
export YARN_USER_CLASSPATH=/opt/Bigdata/client/test/conf/
kinit test
```

Enter the password as prompted. If no error message is displayed, Kerberos authentication is complete.

Step 5 Run the following commands to import data to the HDFS:

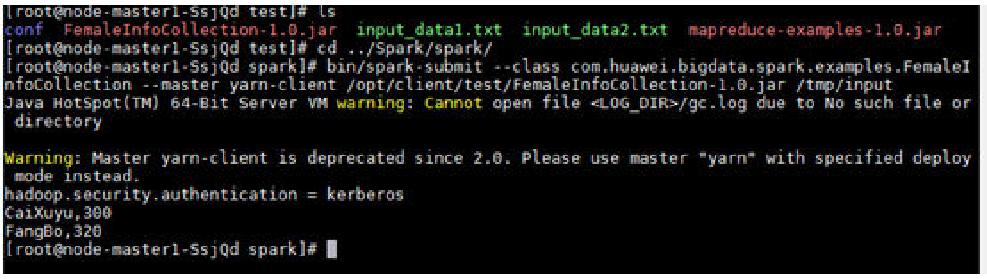
```
cd test
hdfs dfs -mkdir /tmp/input
hdfs dfs -put input_data* /tmp/input
```

Step 6 Run the following commands to run the program:

```
cd /opt/Bigdata/client/Spark/spark
bin/spark-submit --class com.huawei.bigdata.spark.examples.FemaleInfoCollection --master yarn-client /opt/
Bigdata/client/test/FemaleInfoCollection-1.0.jar /tmp/input
```

Step 7 After the program is run successfully, the following information is displayed.

Figure 13-2 Program running result



```
[root@node-master1-SsjQd test]# ls
conf  FemaleInfoCollection-1.0.jar  input_data1.txt  input_data2.txt  mapreduce-examples-1.0.jar
[root@node-master1-SsjQd test]# cd ../Spark/spark/
[root@node-master1-SsjQd spark]# bin/spark-submit --class com.huawei.bigdata.spark.examples.FemaleI
nfoCollection --master yarn-client /opt/client/test/FemaleInfoCollection-1.0.jar /tmp/input
Java HotSpot(TM) 64-Bit Server VM warning: Cannot open file <LOG_DIR>/gc.log due to No such file or
 directory

Warning: Master yarn-client is deprecated since 2.0. Please use master "yarn" with specified deploy
 mode instead.
hadoop.security.authentication = kerberos
CaiXuyu,300
FangBo,320
[root@node-master1-SsjQd spark]# 
```

----End

Running a Hive Program

This section describes how to run a Hive program in security cluster mode.

Prerequisites

You have compiled the program and prepared data files, for example, **hive-examples-1.0.jar**, **input_data1.txt**, and **input_data2.txt**..

Procedure

Step 1 Use a remote login software (for example, MobaXterm) to log in to the master node of the security cluster using SSH (using the EIP).

Step 2 After the login is successful, run the following commands to create the **test** folder in the **/opt/Bigdata/client** directory and create the **conf** folder in the **test** directory:

```
cd /opt/Bigdata/client
mkdir test
cd test
mkdir conf
```

Step 3 Use an upload tool (for example, WinSCP) to copy **FemaleInfoCollection.jar**, **input_data1.txt**, and **input_data2.txt** to the **test** directory, and copy the **keytab**

and **krb5.conf** files obtained in **Step 5** in section **Creating Roles and Users** to the **conf** directory.

Step 4 Run the following commands to configure environment variables and authenticate the created user, for example, **test**:

```
cd /opt/Bigdata/client
source bigdata_env
export YARN_USER_CLASSPATH=/opt/Bigdata/client/test/conf/
kinit test
```

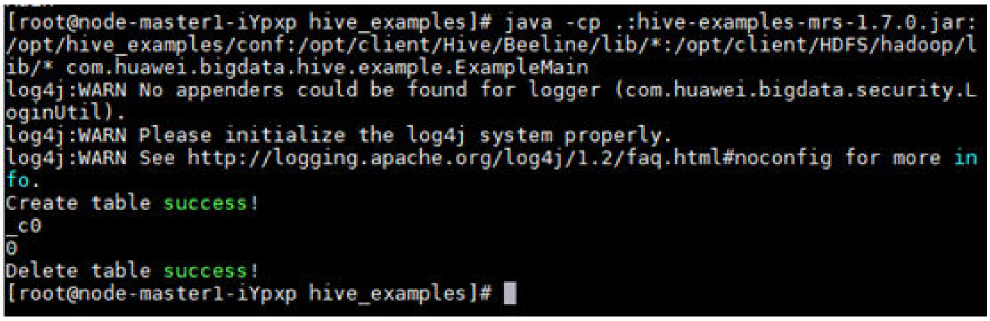
Enter the password as prompted. If no error message is displayed, Kerberos authentication is complete.

Step 5 Run the following command to run the program:

```
chmod +x /opt/hive_examples -R   cd /opt/hive_examples   java -cp .:hive-examples-1.0.jar:/opt/
hive_examples/conf:/opt/Bigdata/client/Hive/Beeline/lib/*:/opt/Bigdata/client/HDFS/hadoop/lib/*
com.huawei.bigdata.hive.example.ExampleMain
```

Step 6 After the program is run successfully, the following information is displayed.

Figure 13-3 Program running result



```
[root@node-master1-iYpxp hive_examples]# java -cp .:hive-examples-mrs-1.7.0.jar:
/opt/hive_examples/conf:/opt/client/Hive/Beeline/lib/*:/opt/client/HDFS/hadoop/l
ib/* com.huawei.bigdata.hive.example.ExampleMain
log4j:WARN No appenders could be found for logger (com.huawei.bigdata.security.L
oginUtil).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more in
fo.
Create table success!
_c0
0
Delete table success!
[root@node-master1-iYpxp hive_examples]#
```

----End

13.6 Terminating a Cluster

You can terminate an MRS cluster that is no longer use after job execution is complete.

Background

You can manually terminate a cluster after data analysis is complete or when the cluster encounters an exception. A cluster failed to be deployed will be automatically terminated.

Procedure

Step 1 Log in to the MRS management console.

Step 2 In the navigation pane on the left, choose **Clusters** > **Active Clusters**.

Step 3 In the cluster list, locate the row containing the cluster to be terminated, and click **Terminate** in the **Operation** column.

The cluster status changes from **Running** to **Terminating**, and finally to **Terminated**. You can view the terminated cluster in **Cluster History**.

----End

14 FAQ

Released On	What's New
2021-06-30	This issue is the fifth official release. Released MRS 3.1.0-LTS.1. For details, see **Creating a Custom Cluster**.
2020-10-24	This issue is the fourth official release. Added the following sections: ● **Methods of Creating MRS Clusters** ● **Quick Creation of a Hadoop Analysis Cluster** ● **Quick Creation of an HBase Analysis Cluster** ● **Quick Creation of a Kafka Streaming Cluster** Modified the following sections: ● **Creating a Custom Cluster** ● **Viewing Basic Cluster Information** ● **Viewing Information of a Historical Cluster**
2020-01-17	This issue is the third official release. Modified the following sections: ● **Creating a Custom Cluster** ● **Viewing Basic Cluster Information** ● **Viewing Information of a Historical Cluster**

Released On	What's New
2019-03-20	This issue is the second official issue. Added the following sections: ● **Security Configuration Suggestions for Clusters with Kerberos Authentication Disabled** ● **Authorizing O&M** ● **Authorizing O&M** ● **Adding a Tag to a Cluster** ● **Installing Third-Party Software Using Bootstrap Actions** ● **Restoring Patches for the Isolated Hosts** ● **Rolling Restart** Modified the following sections: ● **Creating a Custom Cluster** ● **Viewing Basic Cluster Information** ● **Configuring an Auto Scaling Rule** ● **Viewing Information of a Historical Cluster**
2018-10-12	This issue is the first official release.

14.1 MRS Overview

14.1.1 What Is MRS Used For?

MapReduce Service (MRS) is an enterprise-grade big data platform that allows you to quickly build and operate economical, secure, full-stack, cloud-native big data environments on the cloud. It provides engines such as ClickHouse, Spark, Flink, Kafka, and HBase, and supports convergence of data lake, data warehouse, business intelligence (BI), and artificial intelligence (AI). Fully compatible with open-source components, MRS helps you rapidly innovate and expand service growth.

14.1.2 What Types of Distributed Storage Does MRS Support?

MRS supports Hadoop 3.1.*x* and will soon support other mainstream Hadoop versions released by the community. **Table 14-1** lists the component versions supported by MRS.

Table 14-1 MRS component versions

Component	MRS 1.9.2 (Applicable to MRS 1.9.*x*)	MRS 3.1.0
Alluxio	2.0.1	N/A

Component	MRS 1.9.2 (Applicable to MRS 1.9.*x*)	MRS 3.1.0
CarbonData	1.6.1	2.0.1
DBService	1.0.0	2.7.0
Flink	1.7.0	1.12.0
Flume	1.6.0	1.9.0
HBase	1.3.1	2.2.3
HDFS	2.8.3	3.1.1
Hive	2.3.3	3.1.0
Hudi	N/A	0.7.0
Hue	3.11.0	4.7.0
Impala	N/A	3.4.0
Kafka	1.1.0	2.11-2.4.0
KafkaManager	1.3.3.1	N/A
KrbServer	1.15.2	1.17
Kudu	N/A	1.12.1
LdapServer	1.0.0	2.7.0
Loader	2.0.0	N/A
MapReduce	2.8.3	3.1.1
Oozie	N/A	5.1.0
Opentsdb	2.3.0	N/A
Presto	0.216	333
Phoenix (integrated with HBase)	N/A	5.0.0
Ranger	1.0.1	2.0.0
Spark	2.2.2	N/A
Spark2x	N/A	2.4.5
Sqoop	N/A	1.4.7
Storm	1.2.1	N/A
Tez	0.9.1	0.9.2
YARN	2.8.3	3.1.1
ZooKeeper	3.5.1	3.5.6

Component	MRS 1.9.2 (Applicable to MRS 1.9.*x*)	MRS 3.1.0
MRS Manager	1.9.2	N/A
FusionInsight Manager	N/A	8.1.0

14.1.3 How Do I Create an MRS Cluster Using a Custom Security Group?

If you want to use a self-defined security group when buying a cluster, you need to enable port 9022 or select **Auto create** in **Security Group** on the MRS console.

14.1.4 How Do I Use MRS?

MapReduce Service (MRS) is a service you can use to deploy and manage Hadoop-based components on the Cloud. It enables you to deploy Hadoop clusters with a few clicks. MRS provides enterprise-ready big data clusters in the cloud. Tenants can fully control the clusters and easily run big data components such as Hadoop, Spark, HBase, Kafka, and Storm in the clusters.

MRS is easy to use. You can execute various tasks and process or store PB-scale data using computers connected in a cluster. To use MRS, do as follows:

1. Upload local programs and data files to OBS.
2. Create a cluster. You need to specify the cluster type (for example, analysis or streaming), and set ECS instance specifications, number of instances, data disk type (common I/O, high I/O, and ultra-high I/O), and components to be installed, such as Hadoop, Spark, HBase, Hive, Kafka, and Storm, in a cluster. You can use a bootstrap action to install third-party software or modify the cluster running environment on a node before or after the cluster is started.
3. Use MRS to submit, execute, and monitor your programs.
4. Manage clusters on MRS Manager, an enterprise-level unified management platform of big data clusters. You can learn about the health status of services and hosts, obtain critical system information in a timely manner from graphical metric monitoring and customization, modify service attributes based on performance requirements, and start or stop clusters, services, and role instances.
5. Terminate any MRS cluster that you do not require after job execution is complete.

14.1.5 How Does MRS Ensure Security of Data and Services?

MRS is a big data management and analytic platform featuring high security. It ensures data and service security from the following aspects:

- Network isolation

  The public cloud network is divided into service plane and management plane. The two planes are physically isolated to ensure network security.

  - Service plane: provides a network plane for running cluster components. The service plane provides service channels, and implements data access and storage, job submission, and computing.
  - Management plane: provides a public cloud management console that you can use to purchase and manage MRS.
- Host security

  You can deploy third-party antivirus software based on service requirements. MRS provides the following measures to improve security of OSs and ports:

  - OS kernel security hardening
  - OS patch update
  - OS permission control
  - OS port management
  - OS protocol and port attack defense
- Data security

  MRS enables data storage on OBS, thereby ensuring data security.
- Data integrity

  MRS transmits the processed data to OBS using SSL, thereby ensuring data integrity.

14.1.6 Can I Configure a Phoenix Connection Pool?

Phoenix does not support connection pool configuration. You are advised to write code to implement a tool class for managing connections and simulate a connection pool.

14.1.7 Does MRS Support Change of the Network Segment?

You can change the network segment. On the cluster **Dashboard** page of MRS console, click **Change Subnet** to the right of **Default Subnet**, and select a subnet in the VPC of the cluster to expand subnet IP addresses. Selecting a new subnet will not change the IP addresses and subnets of existing nodes.

14.1.8 Can I Downgrade the Specifications of an MRS Cluster Node?

You cannot downgrade the specifications of an MRS cluster node by using the console. If you want to downgrade an MRS cluster node's specifications, contact technical support.

14.1.9 What Is the Relationship Between Hive and Other Components?

- Hive and HDFS

  Hive is an Apache Hadoop project. Hive uses Hadoop Distributed File System (HDFS) as its file storage system. Hive parses and processes structured data stored on HDFS. All data files in the Hive database are stored in HDFS, and all data operations on Hive are also performed using HDFS APIs.
- Hive and MapReduce

All data computing of Hive depends on MapReduce. MapReduce, also an Apache Hadoop project, is a parallel computing framework based on HDFS. During data analysis, Hive parses HiveQL statements submitted by users into MapReduce tasks and submits the tasks for MapReduce to execute.

- Hive and DBService

  MetaStore (metadata service) of Hive processes the structure and attribute information about Hive databases, tables, and partitions that are stored in a relational database. In MRS, the relational database is maintained by DBService.

- Hive and Spark

  Hive data computing can also be implemented on Spark. Spark, also an Apache project, is an in-memory distributed computing framework. During data analysis, Hive parses HiveQL statements submitted by users into Spark tasks and submits the tasks for Spark to execute.

14.1.10 Does an MRS Cluster Support Hive on Spark?

- Clusters of MRS 1.9.*x* support Hive on Spark.
- Clusters of MRS 3.*x* or later support Hive on Spark.
- You can use Hive on Tez for the clusters of other versions.

14.1.11 What Are the Differences Between Hive Versions?

Hive 3.1 has the following differences when compared with Hive 1.2:

- String cannot be converted to int.
- The user-defined functions (UDFs) of the **Date** type are changed to Hive built-in UDFs.
- Hive 3.1 does not provide the index function anymore.
- Hive 3.1 uses the UTC time in time functions, while Hive 1.2 uses the local time zone.
- The JDBC drivers in Hive 3.1 and Hive 1.2 are incompatible.
- In Hive 3.1, column names in ORC files are case-sensitive and underscores-sensitive.
- Hive 3.1 does not allow columns named **time**.

14.1.12 Which MRS Cluster Version Supports Hive Connection and User Synchronization?

MRS cluster 2.0.5 or later supports Hive connections on DataLake Governance Center (DGC) and provides the IAM user synchronization function.

14.1.13 What Are the Differences Between OBS and HDFS in Data Storage?

The data processed by MRS is from OBS or HDFS. OBS is an object-based storage service that provides secure, reliable, and cost-effective storage of huge amounts of data. MRS can directly process data in OBS. You can view, manage, and use data by using the OBS console or OBS client. In addition, you can use REST APIs independently or integrate APIs to service applications to manage and access data.

- Data stored in OBS: Data storage is decoupled from compute. The cluster storage cost is low, and storage capacity is not limited. Clusters can be deleted at any time. However, the computing performance depends on the OBS access performance and is lower than that of HDFS. OBS is recommended for applications that do not demand a lot of computation.
- Data stored in HDFS: Data storage is not decoupled from compute. The cluster storage cost is high, and storage capacity is limited. The computing performance is high. You must export data before you delete clusters. HDFS is recommended for computing-intensive scenarios.

14.1.14 How Do I Obtain the Hadoop Pressure Test Tool?

Download it from **https://github.com/Intel-bigdata/HiBench**.

14.1.15 What Is the Relationship Between Impala and Other Components?

- Impala and HDFS

  Impala uses HDFS as its file storage system. Impala parses and processes structured data, while HDFS provides reliable underlying storage. Impala provides fast data access without moving data in HDFS.
- Impala and Hive

  Impala uses Hive metadata, Open Database Connectivity (ODBC) driver, and SQL syntax. Unlike Hive, which is over MapReduce, Impala implements a distributed architecture based on daemon and handles all query executions on the same node. Therefore, Impala is faster than Hive by reducing the latency caused by MapReduce.
- Impala and MapReduce

  None
- Impala and Spark

  None
- Impala and Kudu

  Kudu can be closely integrated with Impala to replace the combination of Impala, HDFS, and Parquet. You can insert, query, update, and delete data in Kudu tablets using Impala's SQL syntax. In addition, you can use JDBC or ODBC to connect to Kudu for data operations, using Impala as the broker.
- Impala and HBase

  The default Impala tables use data files stored in HDFS, which is ideal for batch loading and query of full table scanning. However, HBase provides convenient and efficient query of OLTP-style organization data.

14.1.16 Statement About the Public IP Addresses in the Open-Source Third-Party SDK Integrated by MRS

The open-source third-party packages on which the open-source components integrated by MRS depend contain SDK usage examples. Public IP addresses such as 12.1.2.3, 54.123.4.56, 203.0.113.0, and 203.0.113.12 are example IP addresses. MRS will not initiate a connection to the public IP address or exchange data with the public IP address.

14.1.17 What Is the Relationship Between Kudu and HBase?

Kudu is designed based on the HBase structure and can implement fast random read/write and update functions that HBase is good at. Kudu and HBase are similar in architecture. The differences are as follows:

- HBase uses ZooKeeper to ensure data consistency, whereas Kudu uses the Raft consensus algorithm to ensure consistency.
- HBase uses HDFS for resilient data storage, whereas Kudu uses TServer to ensure strong data consistency and reliability.

14.1.18 Does MRS Support Running Hive on Kudu?

MRS does not support Hive on Kudu.

Currently, MRS supports only the following two methods to access Kudu:

- Access Kudu through Impala tables.
- Access and operate Kudu tables using the client application.

14.1.19 What Are the Solutions for processing 1 Billion Data Records?

- GaussDB (for MySQL) is recommended for scenarios, such as data updates, online transaction processing (OLTP), and complex analysis of 1 billion data records.
- Impala and Kudu in MRS also meet this requirement. Impala and Kudu can load all join tables to the memory in the join operation.

14.1.20 Can I Change the IP address of DBService?

MRS does not support the change of the DBService IP address.

14.1.21 Can I Clear MRS sudo Logs?

MRS sudo log files record operations performed by user **omm** and are helpful for fault locating. You can delete the logs of the earliest date to release storage space.

1. If the log file is large, add the log file directory to **/etc/logrotate.d/syslog** to enable the system to periodically delete logs.

   Method: Run **sed -i '3 a/var/log/sudo/sudo.log' /etc/logrotate.d/syslog**.
2. Set the maximum number and size of logs in **/etc/logrotate.d/syslog**. If the number or size of logs exceeds the threshold, the logs will be automatically deleted. By default, logs are aged based on the size and number of archived logs. You can use **size** and **rotate** to limit the size and number of archived logs, respectively. If required, you can also add **daily**/**weekly**/**monthly** to specify how often the logs are cleared.

14.1.22 Is the Storm Log also limited to 20 GB in MRS cluster 2.1.0?

In MRS cluster 2.1.0, the Storm log cannot exceed 20 GB. If the Storm log exceeds 20 GB, the log files will be deleted cyclically. Logs are stored on the system disk, therefore, the log space is limited. If you want to keep the log for longer time, mount the log directory to storage media.

14.1.23 What Is Spark ThriftServer?

ThriftServer is a JDBC API. You can use JDBC to connect to ThriftServer to access SparkSQL data. Therefore, you can see JDBCServer in Spark components, but not ThriftServer.

14.1.24 What Access Protocols Are Supported by Kafka?

Kafka supports PLAINTEXT, SSL, SASL_PLAINTEXT, and SASL_SSL.

14.1.25 What Is the Compression Ratio of zstd?

Zstandard (zstd) is an open-source fast lossless compression algorithm. The compression ratio of zstd is twice that of orc. For details, see **https://github.com/L-Angel/compress-demo**. CarbonData does not support lzo, and MRS has zstd integrated.

14.1.26 Why Are the HDFS, YARN, and MapReduce Components Unavailable When an MRS Cluster Is ?

The HDFS, YARN, and MapReduce components are integrated in Hadoop. If the three components are unavailable when are MRS cluster is , select Hadoop instead. After an MRS cluster is created, HDFS, YARN, and MapReduce are available in the **Components** page.

14.1.27 Why Is the ZooKeeper Component Unavailable When an MRS Cluster Is ?

If you create a cluster of a version earlier than MRS 3.*x*, ZooKeeper is installed by default and is not displayed on the GUI.

If you create a cluster of MRS 3.*x* or later, ZooKeeper is available on the GUI and is selected by default.

After the cluster is created, the ZooKeeper component is available on the **Components** page.

14.1.28 Which Python Versions Are Supported by Spark Tasks in an MRS 3.1.0 Cluster?

For MRS 3.1.0 clusters, Python 2.7 or 3.*x* is recommended for Spark tasks.

14.1.29 How Do I Enable Different Service Programs to Use Different YARN Queues?

Create a tenant on Manager.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Tenant Resources**.

Step 2 In the tenant list on the left, select a parent tenant and click ⊕. On the page for adding a sub-tenant, set attributes for the sub-tenant according to **Table 14-2**.

Table 14-2 Sub-tenant parameters

Parameter	Description
Cluster	Indicates the cluster to which the parent tenant belongs.
Parent Tenant Resource	Indicates the name of the parent tenant.
Name	● Indicates the name of the current tenant. The value consists of 3 to 50 characters, including digits, letters, and underscores (_). ● Plan a sub-tenant name based on service requirements. The name cannot be the same as that of a role, HDFS directory, or Yarn queue that exists in the current cluster.
Tenant Type	Specifies whether the tenant is a leaf tenant. ● When **Leaf Tenant** is selected, the current tenant is a leaf tenant and no sub-tenant can be added. ● When **Non-leaf Tenant** is selected, the current tenant is not a leaf tenant and sub-tenants can be added to the current tenant. However, the tenant depth cannot exceed 5 levels.
Computing Resource	Specifies the dynamic computing resources for the current tenant. ● When **Yarn** is selected, the system automatically creates a queue in Yarn and the queue is named the same as the sub-tenant name. – A leaf tenant can directly submit jobs to the queue. – A non-leaf tenant cannot directly submit jobs to the queue. However, Yarn adds an extra queue (hidden) named **default** for the non-leaf tenant to record the remaining resource capacity of the tenant. Actual jobs do not run in this queue. ● If **Yarn** is not selected, the system does not automatically create a queue.
Default Resource Pool Capacity (%)	Indicates the percentage of computing resources used by the current tenant. The base value is the total resources of the parent tenant.

Parameter	Description
Default Resource Pool Max Capacity (%)	Indicates the maximum percentage of computing resources used by the current tenant. The base value is the total resources of the parent tenant.
Storage Resource	Specifies storage resources for the current tenant. ● When **HDFS** is selected, the system automatically creates a folder named after the sub-tenant in the HDFS parent tenant directory. ● When **HDFS** is not selected, the system does not automatically allocate storage resources.
Quota	Indicates the quota for files and directories.
Space Quota	Indicates the quota for the HDFS storage space used by the current tenant. ● If the unit is set to **MB**, the value ranges from **1** to **8796093022208**. If the unit is set to **GB**, the value ranges from **1** to **8589934592**. ● This parameter indicates the maximum HDFS storage space that can be used by the tenant, but not the actual space used. ● If its value is greater than the size of the HDFS physical disk, the maximum space available is the full space of the HDFS physical disk. ● If this quota is greater than the quota of the parent tenant, the actual storage space does not exceed the quota of the parent tenant.
Storage Path	Indicates the HDFS storage directory for the tenant. ● The system automatically creates a folder named after the sub-tenant name in the directory of the parent tenant by default. For example, if the sub-tenant is **ta1s** and the parent directory is **/tenant/ta1**, the storage path for the sub-tenant is then **/tenant/ta1/ta1s**. ● The storage path is customizable in the parent directory.
Description	Indicates the description of the current tenant.

NOTE

Roles, computing resources, and storage resources are automatically created when tenants are created.

- The new role has permissions on the computing and storage resources. This role and its permissions are automatically controlled by the system and cannot be manually managed by choosing **System** > **Permission** > **Role**. The role name is in the format of *Tenant name_Cluster ID*. The ID of the first cluster is not displayed by default.
- When using this tenant, create a system user and bind the user to the role of the tenant. For details, see **Adding a User and Binding the User to a Tenant Role**.
- The sub-tenant can further allocate the resources of its parent tenant. The sum of the resource percentages of direct sub-tenants under a parent tenant at each level cannot exceed 100%. The sum of the computing resource percentages of all level-1 tenants cannot exceed 100%.

Step 3 Check whether the current tenant needs to be associated with resources of other services.

- If yes, go to **Step 4**.
- If no, go to **Step 5**.

Step 4 Click **Associate Service** to configure other service resources used by the current tenant.

1. Set **Services** to **HBase**.
2. Set **Association Type** as follows:
   - **Exclusive** indicates that the service resources are used by the tenant exclusively and cannot be associated with other tenants.
   - **Shared** indicates that the service resources can be shared with other tenants.

NOTE

- Only HBase can be associated with a new tenant. However, HDFS, HBase, and Yarn can be associated with existing tenants.
- To associate an existing tenant with service resources, click the target tenant in the tenant list, switch to the **Service Associations** page, and click **Associate Service** to configure resources to be associated with the tenant.
- To disassociate an existing tenant from service resources, click the target tenant in the tenant list, switch to the **Service Associations** page, and click **Delete** in the **Operation** column. In the displayed dialog box, select **I have read the information and understand the impact** and click **OK**.

3. Click **OK**.

Step 5 Click **OK**. Wait until the system displays a message indicating that the tenant is successfully created.

----End

14.1.30 Differences and Relationships Between the MRS Management Console and Cluster Manager

You can access Manager from the MRS management console.

Manager is classified as MRS Manager and FusionInsight Manager.

- MRS Manager is the manager page of MRS 2.*x* or earlier clusters.
- FusionInsight Manager is the manager page of MRS 3.*x* or later clusters.

The following table lists the differences and relationships between the management console and FusionInsight Manager.

Common Operation	MRS Console	FusionInsight Manager
Changing subnets, adding security group rules, controlling OBS permissions, managing agencies, and synchronizing IAM users	Supported	Not supported
Adding node groups, scaling out, scaling in, and upgrading specifications	Supported	Not supported
Isolating hosts, starting all roles, and stopping all roles	Supported	Supported
Downloading the client, starting services, stopping services, and perform rolling restart of services	Supported	Supported
Viewing the instance status of services, configuring parameters, and synchronizing configurations	Supported	Supported
Viewing cleared alarms and events	Supported	Supported
Viewing the alarm help	Not supported	Supported
Setting thresholds	Not supported	Supported
Adding message subscription specifications	Supported	Not supported
Managing files	Supported	Not supported
Managing jobs	Supported	Not supported
Managing tenants	Supported	Supported
Managing tags	Supported	Not supported

Common Operation	MRS Console	FusionInsight Manager
Managing permissions (adding and deleting users, user groups, and roles, and changing passwords)	Not supported	Supported
Performing backup and restoration	Not supported	Supported
Auditing	Not supported	Supported
Monitoring resources and logging	Supported	Supported

14.1.31 How Do I Unbind an EIP from an MRS Cluster Node?

Symptom

After an EIP is bound on the console, the EIP cannot be unbound in the EIP module of the VPC service.

A dialog box is displayed, indicating that the operation cannot be performed because the EIP is being used by MapReduce.

Procedure

Step 1 Log in to the VPC console and choose **Virtual Private Cloud** > **My VPCs**. Find the target VPC in the VPC list.

Step 2 Click the VPC name to go to the **Summary** tab page and click the number next to **Subnets** in the **Networking Components** area to find the subnet to which the cluster belongs.

Step 3 In the subnet list, click the target subnet name. Click the **IP Addresses** tab, locate the target public IP address and click **Unbind from EIP** in the **Operation** column.

----End

14.2 Account and Password

14.2.1 What Is the Account for Logging In to Manager?

The default account for logging in to Manager is **admin**, and the password is the one you set when you created the cluster.

14.2.2 How Do I Query and Change the Password Validity Period of an Account?

Querying the Password Validity Period

Querying the password validity period of a component running user (human-machine user or machine-machine user):

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to switch to the client directory, for example, **/opt/Bigdata/client**:

cd /opt/Bigdata/client

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 Run the following command and enter the password of user **kadmin/admin** to log in to the kadmin console:

kadmin -p kadmin/admin

NOTE

The default password of user **kadmin/admin** is **Admin@123**. Change the password upon your first login or as prompted and keep the new password secure.

Step 5 Run the following command to view the user information:

getprinc *Internal system username*

Example: **getprinc user1**

```
kadmin:  getprinc user1
......
Expiration date: [never]
Last password change: Sun Oct 09 15:29:54 CST 2022
Password expiration date: [never]
......
```

----End

Querying the password validity period of an OS user:

Step 1 Log in to any master node in the cluster as user **root**.

Step 2 Run the following command to view the password validity period (value of **Password expires**):

chage -l *Username*

For example, to view the password validity period of user **root**, run the **chage -l root** command. The command output is as follows:

```
[root@xxx ~]#chage -l root
Last password change                                    : Sep 12, 2021
Password expires                                        : never
Password inactive                                       : never
Account expires                                         : never
Minimum number of days between password change          : 0
```

```
Maximum number of days between password change         : 99999
Number of days of warning before password expires       : 7
```

----End

Changing the Password Validity Period

- The password of a machine-machine user is randomly generated and never expires by default.
- The password validity period of a human-machine user can be changed by modifying the password policy on Manager.

14.3 Accounts and Permissions

14.3.1 Does an MRS Cluster Support Access Permission Control If Kerberos Authentication Is not Enabled?

For MRS cluster 2.1.0 or earlier, choose **System** > **Configuration** > **Permission** on MRS Manager.

For MRS cluster 3.*x* or later, choose **System** > **Permission** on FusionInsight Manager.

14.3.2 How Do I Assign Tenant Management Permission to a New Account?

You can assign tenant management permission only in analysis or hybrid clusters, but not in streaming clusters.

The operations vary depending on the MRS cluster version:

Procedure for versions earlier than MRS cluster 3.x:

Step 1 Log in to MRS Manager as user **admin**.

Step 2 Choose **System** > **Manage User**. Select the new account, and click **Modify** in the **Operation** column.

Step 3 In **Assign Rights by Role**, click **Select and Add Role**.

- If you bind the **Manager_tenant** role to the account, the account will have permission to view tenant management information.
- If you bind the **Manager_administrator** role to the account, the account will have permission to view and perform tenant management.

Step 4 Click **OK**.

----End

Procedure for MRS cluster 3.x and later versions:

Step 1 Log in to FusionInsight Manager and choose **System** > **Permission** > **User**.

Step 2 Locate the user and click **Modify**.

Modify the parameters based on service requirements.

If you bind the **Manager_tenant** role to the account, the account will have permission to view tenant management information. If you bind the **Manager_administrator** role to the account, the account will have permission to perform tenant management and view related information.

NOTE

It takes about three minutes for the settings to take effect after user group or role permission are modified.

Step 3 Click **OK**.

----End

14.3.3 How Do I Customize an MRS Policy?

1. On the IAM console, choose **Permissions** in the navigation pane, and click **Create Custom Policy**.
2. Set a policy name in **Policy Name**.
3. Set **Scope** to **Project-level service** for MRS.
4. Specify **Policy View**. The following options are supported:
   - **Visual editor**: Select cloud services, actions, resources, and request conditions from the navigation pane to customize the policy. You do not require knowledge of JSON syntax.
   - **JSON**: Edit JSON policies from scratch or based on an existing policy.

   You can also click **Select Existing Policy/Role** in the **Policy Content** area to select an existing policy as the template for modification.
5. (Optional) Enter a brief description in the **Description** area.
6. Click **OK**.
7. Attach the policy to a user group. Users in the group then inherit the permissions defined in this policy.

14.3.4 Why Is the Manage User Function Unavailable on the System Page on MRS Manager?

Check whether you have the **Manager_administrator** permission. If you do not have this permission, **Manage User** will not be available on the **System** page of MRS Manager.

14.3.5 Does Hue Support Account Permission Configuration?

Hue does not provide an entry for configuring account permissions on its web UI. However, you can configure user roles and user groups for Hue accounts on the **System** tab on Manager.

14.4 Client Usage

14.4.1 How Do I Configure Environment Variables and Run Commands on a Component Client?

1. Log in to any Master node as user **root**.
2. Run the **su - omm** command to switch to user **omm**.
3. Run the **cd /opt/client** command to switch to the client.
4. Run the **source bigdata_env** command to configure environment variables.

   If Kerberos authentication is enabled for the current cluster, run the **kinit** *Component service user* command to authenticate the user. If Kerberos authentication is disabled, skip this step.
5. After the environment variables are configured, run the client command of the component. For example, to view component information, you can run the HDFS client command **hdfs dfs -ls /** to view the HDFS root directory file.

14.4.2 How Do I Disable ZooKeeper SASL Authentication?

Log in to FusionInsight Manager, choose **Cluster** > **Services** > **ZooKeeper**, click the **Configurations** tab and then **All Configurations**. In the navigation pane on the left, choose **quorumpeer(Role)** > **Customization**, add the **set zookeeper.sasl.disable** parameter, and set its value to **false**. Save the configuration and restart the ZooKeeper service.

14.4.3 An Error Is Reported When the kinit Command Is Executed on a Client Node Outside an MRS Cluster

Symptom

After the client is installed on a node outside an MRS cluster and the **kinit** command is executed, the following error information is displayed:

```
-bash kinit Permission denied
```

The following error information is displayed when the **java** command is executed:

```
-bash: /xxx/java: Permission denied
```

After running the **ll /***Java installation path***/JDK/jdk/bin/java** command, it is found that the file execution permission is correct.

Fault Locating

Run the **mount | column -t** command to check the status of the mounted partition. It is found that the partition status of the mount point where the Java execution file is located is **noexec**. In the current environment, the data disk where the MRS client is installed is set to **noexec**, that is, binary file execution is prohibited. As a result, Java commands cannot be executed.

Solution

1. Log in to the node where the MRS client is located as user **root**.
2. Remove the configuration item **noexec** of the data disk where the MRS client is located from the **/etc/fstab** file.

3. Run the **umount** command to detach the data disk, and then run the **mount -a** command to remount the data disk.

14.5 Web Page Access

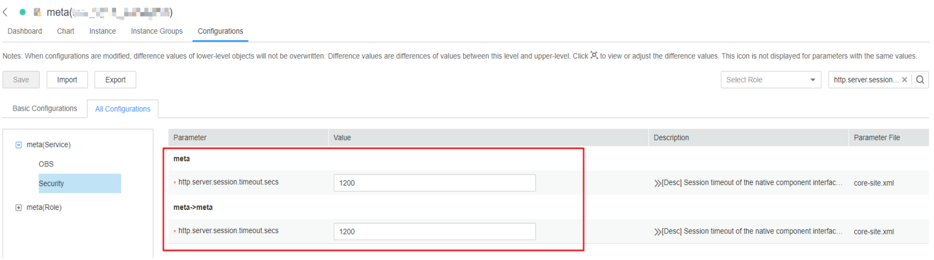
14.5.1 How Do I Change the Session Timeout Duration for an Open Source Component Web UI?

You need to set a proper web session timeout duration for security purposes. To change the session timeout duration, do as follows:

Checking Whether the Cluster Supports Session Timeout Duration Adjustment

- For MRS cluster versions earlier than 3.x:
  a. On the cluster details page, choose **Components** > **meta** > **Service Configuration**.
  b. Switch **Basic** to **All**, and search for the **http.server.session.timeout.secs**.

     If **http.server.session.timeout.secs** does not exist, the cluster does not support change of the session timeout duration. If the parameter exists, perform the following steps to modify it.
- MRS 3.x and later: Log in to FusionInsight Manager and choose **Cluster** > **Services** > **meta**. On the displayed page, click **Configurations** and select **All Configurations**. Search for the **http.server.session.timeout.secs** configuration item. If this configuration item exists, perform the following steps to modify it. If the configuration item does not exist, the version does not support dynamic adjustment of the session duration.



Notes: When configurations are modified, difference values of lower-level objects will not be overwritten. Difference values are differences of values between this level and upper-level. Click to view or adjust the difference values. This icon is not displayed for parameters with the same values.

Parameter	Value	Description	Parameter File
meta			
* http.server.session.timeout.secs	1200	[Desc] Session timeout of the native component interfac...	core-site.xml
meta->meta			
* http.server.session.timeout.secs	1200	[Desc] Session timeout of the native component interfac...	core-site.xml

You are advised to set all session timeout durations to the same value. Otherwise, the settings of some parameters may not take effect due to value conflict.

Modifying the Timeout Duration on Manager and the Authentication Center Page

For clusters of versions earlier than MRS 3.*x*:

1. Log in to each master node in the cluster and perform 2 to 4.
2. Change the value of **<session-timeout>20</session-timeout>** in the **/opt/Bigdata/apache-tomcat-7.0.78/webapps/cas/WEB-INF/web.xml** file. **<session-timeout>20</session-timeout>** indicates the session timeout duration, in minutes. Change it based on service requirements. The maximum value is 480 minutes.

3. Change the value of **<session-timeout>20</session-timeout>** in the **/opt/Bigdata/apache-tomcat-7.0.78/webapps/web/WEB-INF/web.xml** file. **<session-timeout>20</session-timeout>** indicates the session timeout duration, in minutes. Change it based on service requirements. The maximum value is 480 minutes.
4. Change the values of **p:maxTimeToLiveInSeconds="${tgt.maxTimeToLiveInSeconds:1200}"** and **p:timeToKillInSeconds="${tgt.timeToKillInSeconds:1200}"** in the **/opt/Bigdata/apache-tomcat-7.0.78/webapps/cas/WEB-INF/spring-configuration/ticketExpirationPolicies.xml** file. The maximum value is 28,800 seconds.
5. Restart the Tomcat node on the active master node.
   a. On the active master node, run the **netstat -anp |grep 28443 |grep LISTEN | awk '{print $7}'** command as user **omm** to query the Tomcat process ID.
   b. Run the **kill -9** *{pid}* command, in which *{pid}* indicates the Tomcat process ID obtained in **5.a**.
   c. Wait until the process automatically restarts. You can run the **netstat -anp |grep 28443 |grep LISTEN** command to check whether the process is successfully restarted. If the process can be queried, the process is successfully restarted. If the process cannot be queried, query the process again later.

For clusters of MRS 3.*x* or later

1. Log in to each master node in the cluster and perform **2** to **3** on each master node.
2. Change the value of **<session-timeout>20</session-timeout>** in the **/opt/Bigdata/om-server_xxx/apache-tomcat-xxx/webapps/web/WEB-INF/web.xml** file. **<session-timeout>20</session-timeout>** indicates the session timeout duration, in minutes. Change it based on service requirements. The maximum value is 480 minutes.
3. Add **ticket.tgt.timeToKillInSeconds=28800** to the **/opt/Bigdata/om-server_xxx/apache-tomcat-8.5.63/webapps/cas/WEB-INF/classes/config/application.properties** file. **ticket.tgt.timeToKillInSeconds** indicates the validity period of the authentication center, in seconds. Change it based on service requirements. The maximum value is 28,800 seconds.
4. Restart the Tomcat node on the active master node.
   a. On the active master node, run the **netstat -anp |grep 28443 |grep LISTEN | awk '{print $7}'** command as user **omm** to query the Tomcat process ID.
   b. Run the **kill -9** *{pid}* command, in which *{pid}* indicates the Tomcat process ID obtained in **4.a**.
   c. Wait until the process automatically restarts.

      You can run the **netstat -anp |grep 28443 |grep LISTEN** command to check whether the process is successfully restarted. If the process is displayed, the process is successfully restarted. If the process is not displayed, query the process again later.

Modifying the Timeout Duration for an Open-Source Component Web UI

1. Access the **All Configurations** page.
   - For MRS cluster versions earlier than MRS 3.x:

     On the cluster details page, choose **Components > Meta > Service Configuration**.
   - For MRS cluster version 3.*x* or later:

     Log in to FusionInsight Manager and choose **Cluster** > **Services** > **meta**. On the displayed page, click **Configurations** and select **All Configurations**.
2. Change the value of **http.server.session.timeout.secs** under **meta** as required. The unit is second.
3. Save the settings, deselect **Restart the affected services or instances**, and click **OK**.

   You are advised to perform the restart during off-peak hours.
4. (Optional) If you need to use the Spark web UI, search for **spark.session.maxAge** on the **All Configurations** page of Spark and change the value (in seconds).

   Save the settings, deselect **Restart the affected services or instances**, and click **OK**.
5. Restart the meta service and components on web UI, or restart the cluster during off-peak hours.

   To prevent service interruption, restart the service during off-peak hours or perform a rolling restart.

14.5.2 Why Cannot I Refresh the Dynamic Resource Plan Page on MRS Tenant Tab?

Step 1 Log in to the Master1 and Master2 nodes as user **root**.

Step 2 Run the **ps -ef |grep aos** command to check the AOS process ID.

Step 3 Run the **kill -9** *AOS process ID* command to end the AOS process.

Step 4 Wait until the AOS process is automatically restarted.

You can run the **ps -ef |grep aos** command to check whether the AOS process restarts successfully. If the process exists, the restart is successful and the **Dynamic Resource Plan** page will be refreshed. If the process does not exist, retry later.

----End

14.5.3 What Do I Do If the Kafka Topic Monitoring Tab Is Unavailable on Manager?

Step 1 Log in to each Master node of the cluster and switch to user **omm**.

Step 2 Go to the **/opt/Bigdata/apache-tomcat-7.0.78/webapps/web/WEB-INF/lib/components/Kafka/** directory.

Step 3 Run the **cp /opt/share/zookeeper-3.5.1-mrs-2.0/zookeeper-3.5.1-mrs-2.0.jar ./** command to copy the ZooKeeper package.

Step 4 Restart the Tomcat process.

sh /opt/Bigdata/apache-tomcat-7.0.78/bin/shutdown.sh

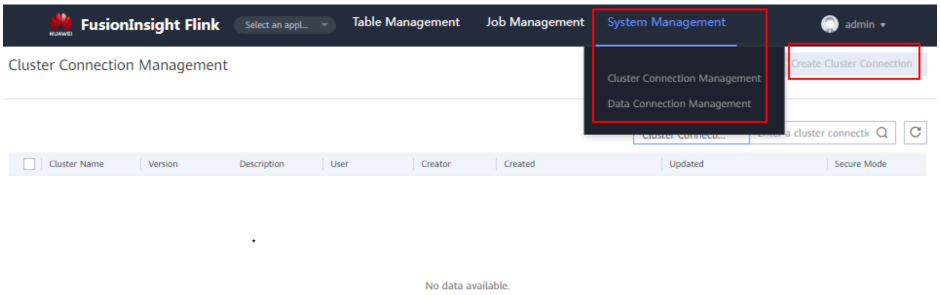
sh /opt/Bigdata/apache-tomcat-7.0.78/bin/startup.sh

----End

14.5.4 How Do I Do If an Error Is Reported or Some Functions Are Unavailable When I Access the Web UIs of HDFS, Hue, YARN, and Flink?

Users who access the web UIs of components such as HDFS, Hue, YARN, and Flink do not have required management permissions. As a result, an error is reported or some functions are unavailable. The following are some examples:

- After you log in to the web UI of Flink as the current user, some content cannot be displayed, and you do not have the permission to create applications, cluster connections, or data connections.

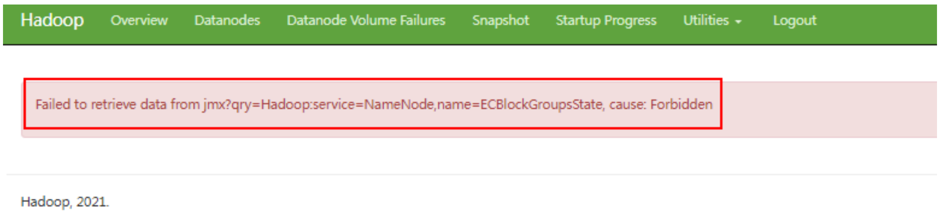


Cluster Connection Management

☐	Cluster Name	Version	Description	User	Creator	Created	Updated	Secure Mode

No data available.

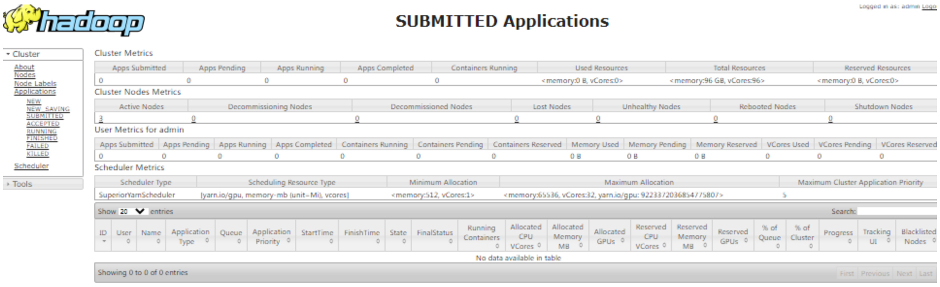
- After you log in to the web UI of HDFS as the current user, error message "Failed to retrieve data from /jmx?qry=java.lang:type=Memory, cause: Forbidden" is displayed.



Failed to retrieve data from jmx?qry=Hadoop:service=NameNode,name=ECBlockGroupsState, cause: Forbidden

Hadoop, 2021.

- After you log in to the web UI of YARN as the current user, you cannot view job information.



SUBMITTED Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	Reserved Resources
0	0	0	0	0	<memory:0 B, vCores:0>	<memory:96 GB, vCores:96>	<memory:0 B, vCores:0>

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
3	0	0	0	0	0	0

User Metrics for admin

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved	VCores Used	VCores Pending	VCores Reserved
0	0	0	0	0	0	0	0 B	0 B	0 B	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
SuperiorYarnScheduler	[yarn.io/gpu, memory-mb (unit=Mi), vcores]	<memory:512, vCores:1>	<memory:65536, vCores:32, yarn.io/gpu: 9223372036854775807>	5

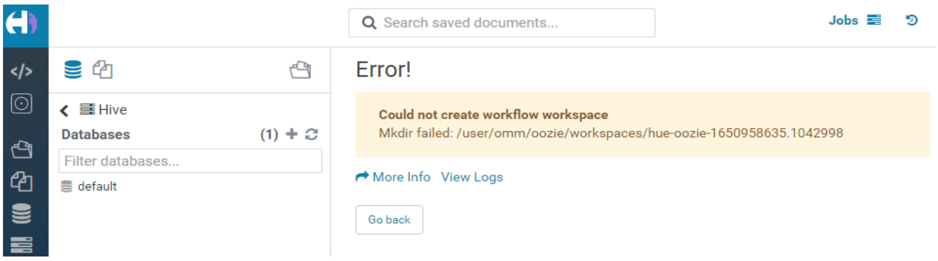
Show 20 entries Search:

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Allocated GPUs	Reserved CPU VCores	Reserved Memory MB	Reserved GPUs	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes

No data available in table

Showing 0 to 0 of 0 entries First Previous Next Last

- After you log in to the web UI of Hue as the current user, click ⊙ in the navigation pane on the left, and select **Workflow**, an error message is displayed.



Could not create workflow workspace
Mkdir failed: /user/omm/oozie/workspaces/hue-oozie-1650958635.1042998

You are advised to log in to the web UIs of the components as a user with corresponding management permissions. For example, you can create a service user who has the management permissions on HDFS and you can log in to the web UI of HDFS as the created user.

14.6 Alarm Monitoring

14.6.1 In an MRS Streaming Cluster, Can the Kafka Topic Monitoring Function Send Alarm Notifications?

The Kafka topic monitoring function cannot send alarms by email or SMS message. However, you can view alarm information on Manager.

14.7 Performance Tuning

14.7.1 Does an MRS Cluster Support System Reinstallation?

An MRS cluster does not support system reinstallation.

14.7.2 Can I Change the OS of an MRS Cluster?

The OS of an MRS cluster cannot be changed.

14.7.3 How Do I Improve the Resource Utilization of Core Nodes in a Cluster?

1. Go to the Yarn service configuration page.
   - For versions earlier than MRS 2.0.1, log in to MRS Manager, choose **Services** > **Yarn** > **Service Configuration**, and select **All** from the **Basic** drop-down list.
   - For MRS 2.0.1 or later, click the cluster name on the MRS console, choose **Components** > **Yarn** > **Service Configuration**, and select **All** from the **Basic** drop-down list.

NOTE

If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

- For MRS 3.x or later, log in to FusionInsight Manager. Choose **Cluster** > *Name of the desired cluster* > **Services** > **Yarn** > **Configurations** > **All Configurations**.

2. Search for **yarn.nodemanager.resource.memory-mb**, and increase the value based on the actual memory of the cluster nodes.
3. Save the change and restart the affected services or instances.

14.7.4 How Do I Stop the Firewall Service?

Step 1 Log in to each node of a cluster as user **root**.

Step 2 Check whether the firewall service is started.

For example, to check the firewall status on EulerOS, run the **systemctl status firewalld.service** command.

Step 3 Stop the firewall service.

For example, to stop the firewall service on EulerOS, run the **systemctl stop firewalld.service** command.

----End

14.8 Job Development

14.8.1 How Do I Get My Data into OBS or HDFS?

MRS can process data in OBS and HDFS. You can get your data into OBS or HDFS as follows:

1. Upload local data to OBS.
   a. Log in to the OBS console.
   b. Create a parallel file system named **userdata** on OBS and create the **program**, **input**, **output**, and **log** folders in the file system.
      i. Choose **Parallel File System** > **Create Parallel File System**, and create a file system named **userdata**.
      ii. In the OBS file system list, click the file system name **userdata**, choose **Files** > **Create Folder**, and create the **program**, **input**, **output**, and **log** folders.
   c. Upload data to the **userdata** file system.
      i. Go to the **program** folder and click **Upload File**.
      ii. Click **add file** and select a user program.
      iii. Click **Upload**.
      iv. Upload the user data file to the **input** directory using the same method.

2. Import OBS data to HDFS.

   You can import OBS data to HDFS only when **Kerberos Authentication** is disabled and the cluster is running.

   a. Log in to the MRS console.
   b. Click the name of the cluster.
   c. On the page displayed, select the **Files** tab page and click **HDFS File List**.
   d. Select a data directory, for example, **bd_app1**.

      The **bd_app1** directory is only an example. You can use any directory on the page or create a new one.

   e. Click **Import Data** and click **Browse** to select an OBS path and an HDFS path.
   f. Click **OK**.

      You can view the file upload progress on the **File Operation Records** tab page.

14.8.2 What Types of Spark Jobs Can Be Submitted in a Cluster?

MRS clusters support Spark jobs submitted in Spark, Spark Script, or Spark SQL mode.

14.8.3 Can I Run Multiple Spark Tasks at the Same Time After the Minimum Tenant Resources of an MRS Cluster Is Changed to 0?

You can run only one Spark task at a time after the minimum tenant resources of an MRS cluster is changed to 0.

14.8.4 What Are the Differences Between the Client Mode and Cluster Mode of Spark Jobs?

You need to understand the concept ApplicationMaster before understanding the essential differences between Yarn-client and Yarn-cluster.

In Yarn, each application instance has an ApplicationMaster process, which is the first container started by the application. It interacts with ResourceManager and requests resources. After obtaining resources, it instructs NodeManager to start containers. The essential difference between the Yarn-cluster and Yarn-client modes lies in the ApplicationMaster process.

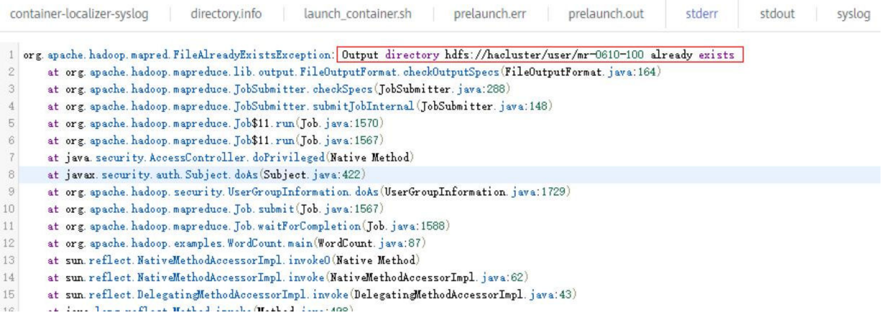
In Yarn-cluster mode, Driver runs in ApplicationMaster, which requests resources from Yarn and monitors the running status of a job. After a user submits a job, the client can be stopped and the job continues running on Yarn. Therefore, the Yarn-cluster mode is not suitable for running interactive jobs.

In Yarn-client mode, ApplicationMaster requests only Executor from Yarn. The client communicates with the requested containers to schedule tasks. Therefore, the client cannot be stopped.

14.8.5 How Do I View MRS Job Logs?

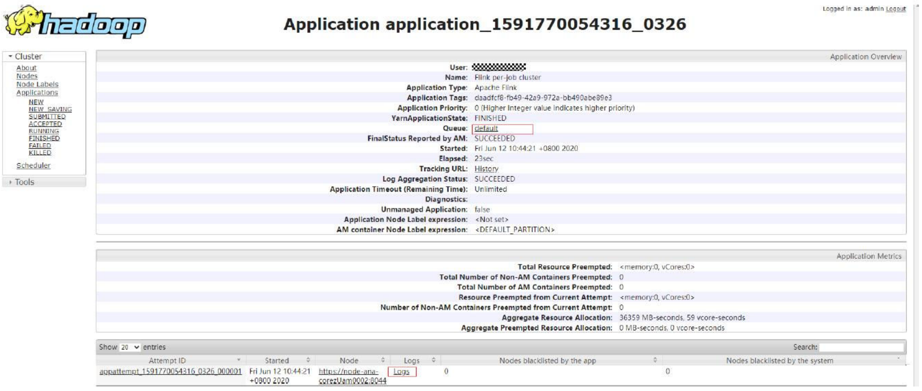
Step 1 On the **Jobs** page of the MRS console, you can view logs of each job, including launcherJob and realJob logs.

- Generally, error logs are printed in **stderr** and **stdout** for launcherJob jobs, as shown in the following figure:



```
org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://hacluster/user/mr-0610-100 already exists
    at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:164)
    at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:288)
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:148)
    at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1570)
    at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1567)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1729)
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1567)
    at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1588)
    at org.apache.hadoop.examples.WordCount.main(WordCount.java:87)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
```

- You can view realJob logs on the ResourceManager web UI provided by the Yarn service on MRS Manager.



Application application_1591770054316_0326

User: [redacted]
Name: Flink per-job cluster
Application Type: Apache Flink
Application Tags: daadfcf8-fb49-42a9-972a-bb490abe89e3
Application Priority: 0 (Higher Integer value indicates higher priority)
YarnApplicationState: FINISHED
Queue: default
FinalStatus Reported by AM: SUCCEEDED
Started: Fri Jun 12 10:44:21 +0800 2020
Elapsed: 23sec
Tracking URL: History
Log Aggregation Status: SUCCEEDED
Application Timeout (Remaining Time): Unlimited
Diagnostics:
Unmanaged Application: false
Application Node Label expression: <Not set>
AM container Node Label expression: <DEFAULT_PARTITION>

Total Resource Preempted: <memory:0, vCores:0>
Total Number of Non-AM Containers Preempted: 0
Total Number of AM Containers Preempted: 0
Resource Preempted from Current Attempt: <memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt: 0
Aggregate Resource Allocation: 36359 MB-seconds, 59 vcore-seconds
Aggregate Preempted Resource Allocation: 0 MB-seconds, 0 vcore-seconds

Attempt ID	Started	Node	Logs	Nodes blacklisted by the app	Nodes blacklisted by the system
appattempt_1591770054316_0326_000001	Fri Jun 12 10:44:21 +0800 2020	https://node-ana-corezUam0002:8044	Logs	0	0

Step 2 Log in to the Master node of the cluster to obtain the job log files in **Step 1**. The HDFS path is **/tmp/logs/***{submit_user}***/logs/***{application_id}*.

Step 3 After the job is submitted, if the job application ID cannot be found on the Yarn web UI, the job fails to be submitted. You can log in to the active Master node of the cluster and view the job submission process log **/var/log/executor/logs/exe.log**.

----End

14.8.6 How Do I Do If the Message "The current user does not exist on MRS Manager. Grant the user sufficient permissions on IAM and then perform IAM user synchronization on the Dashboard tab page." Is Displayed?

If IAM synchronization is not performed when a job is submitted in a security cluster, the error message "The current user does not exist on MRS Manager. Grant the user sufficient permissions on IAM and then perform IAM user synchronization on the Dashboard tab page." is displayed.

Before submitting a job, on the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.

14.8.7 LauncherJob Job Execution Is Failed And the Error Message "jobPropertiesMap is null." Is Displayed

The cause of the launcherJob failure is that the user who submits the job does not have the write permission on the **hdfs /mrs/job-properties** directory.

This problem is fixed in the 2.1.0.6 patch. You can also grant the write permission on the **/mrs/job-properties** directory to the synchronized user who submits the job on MRS Manager.

14.8.8 How Do I Do If the Flink Job Status on the MRS Console Is Inconsistent with That on Yarn?

To save storage space, the Yarn configuration item **yarn.resourcemanager.max-completed-applications** is modified to reduce the number of historical job records stored on Yarn. Flink jobs are long-term jobs. The realJob is still running on Yarn, but the launcherJob has been deleted. As a result, the launcherJob cannot be found on Yarn, and the job status fails to be updated. This problem is fixed in the 2.1.0.6 patch.

Workaround: Terminate the job whose launcherJob cannot be found. The status of the job submitted later will be updated.

14.8.9 How Do I Do If a SparkStreaming Job Fails After Being Executed Dozens of Hours and the OBS Access 403 Error is Reported?

When a user submits a job that needs to read and write OBS, the job submission program adds the temporary access key (AK) and secret key (SK) for accessing OBS by default. However, the temporary AK and SK have expiration time.

If you want to run long-term jobs such as Flink and SparkStreaming, you can enter the AK and SK in **Service Parameter** to ensure that the jobs will not fail to be executed due to key expiration.

14.8.10 How Do I Do If an Alarm Is Reported Indicating that the Memory Is Insufficient When I Execute a SQL Statement on the ClickHouse Client?

Symptom

The ClickHouse client restricts the memory used by GROUP BY statements. When a SQL statement is executed on the ClickHouse client, the following error information is displayed:

```
Progress: 1.83 billion rows, 85.31 GB (68.80 million rows/s., 3.21 GB/s.)        6%Received exception from
server:
Code: 241. DB::Exception: Received from localhost:9000, 127.0.0.1.
DB::Exception: Memory limit (for query) exceeded: would use 9.31 GiB (attempt to allocate chunk of
1048576 bytes), maximum: 9.31 GiB:
(while reading column hits):
```

Solution

- Run the following command before executing an SQL statement on condition that the cluster has sufficient memory:
  ```
  SET max_memory_usage = 128000000000; #128G
  ```
- If no sufficient memory is available, ClickHouse enables you to overflow data to disk to free up the memory: You are advised to set the value of **max_memory_usage** to twice the size of **max_bytes_before_external_group_by**.
  ```
  set max_bytes_before_external_group_by=20000000000; #20G
  set max_memory_usage=40000000000; #40G
  ```

14.8.11 Why Submitted Yarn Job Cannot Be Viewed on the Web UI?

After a Yarn job is created, it cannot be viewed if you log in to the web UI as the **admin** user.

- The **admin** user is a user on the cluster management page. Check whether the user has the **supergroup** permission. Generally, only the user with the **supergroup** permission can view jobs.
- Log in to Yarn as the user who submits jobs to view jobs on Yarn. Do not view the jobs using the **admin** user.

14.8.12 How Do I Modify the HDFS NameSpace (fs.defaultFS) of an Existing Cluster?

You can modify or add the HDFS NameSpace (fs.defaultFS) of the cluster by modifying the **core-site.xml** and **hdfs-site.xml** files on the client. However, you are not advised to perform this operation on the server.

14.8.13 How Do I Do If the launcher-job Queue Is Stopped by YARN due to Insufficient Heap Size When I Submit a Flink Job on the Management Plane?

Symptom

The launcher-job queue is stopped by YARN when a Flink job is submitted on the management plane.

Solution

Increase the heap size of the launcher-job queue.

1. Log in to the active OMS node as user **omm**.
2. Change the value of **job.launcher.resource.memory.mb** in **/opt/executor/webapps/executor/WEB-INF/classes/servicebroker.xml** to **2048**.
3. Run the **sh /opt/executor/bin/restart-executor.sh** command to restart the executor process.

14.9 Cluster Upgrade/Patching

14.9.1 Can I Upgrade an MRS Cluster?

You cannot upgrade an MRS cluster. However, you can create a cluster of the target version and migrate data from the old cluster to the new cluster.

14.9.2 Can I Change the MRS Cluster Version?

You cannot change the version of an MRS cluster. However, you can terminate the current cluster and create an MRS cluster of the version you require.

14.10 Cluster Access

14.10.1 Can I Switch Between the Two Login Modes of MRS?

No. You can select the login mode when creating the cluster. You cannot change the login mode after you created the cluster.

14.10.2 How Can I Obtain the IP Address and Port Number of a ZooKeeper Instance?

You can obtain the IP address and port number of a ZooKeeper instance through the MRS console or FusionInsight Manager.

Method 1: Obtaining the IP address and port number of a ZooKeeper through the MRS console

1. On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.
2. Click the **Components** tab and choose **ZooKeeper**. On the displayed page, click **Instances** to view the business IP address of a ZooKeeper instance.
3. Click the **Service Configuration** tab. On the displayed page, search for the **clientPort** parameter to view the port number of the ZooKeeper instance.

Method 2: Obtaining the IP address and port number of a ZooKeeper through FusionInsight Manager

1. Log in to FusionInsight Manager. For details, see .
2. Perform the following operations to obtain the IP address and port number of a ZooKeeper instance.
   - For clusters of MRS 3.*x* or earlier
     1. Choose **Services** > **ZooKeeper**. On the displayed page, click the **Instance** tab to view the business IP address of a ZooKeeper instance.
     2. Click the **Service Configuration** tab. On the displayed page, search for the **clientPort** parameter to view the port number of the ZooKeeper instance.

– For clusters of MRS 3.*x* or later

i. Choose **Cluster** > **Services** > **ZooKeeper**. On the displayed page, click the **Instance** tab to view the business IP address of a ZooKeeper instance.

ii. Click the **Configurations** tab. On the displayed page, search for the **clientPort** parameter to view the port number of the ZooKeeper instance.

14.10.3 How Do I Do If a New Node Cannot Be logged In to as a Linux User?

If you can log in to an existing node as the Linux user but fail to log in to the newly added node, log in to the newly added node as the root user.

14.10.4 How Do I Access an MRS Cluster from a Node Outside the Cluster?

Creating a Linux ECS Outside the Cluster to Access the MRS Cluster

Step 1 Create an ECS outside the cluster.

Set **AZ**, **VPC**, and **Security Group** of the ECS to the same values as those of the cluster to be accessed.

Step 2 On the VPC management console, apply for an EIP and bind it to the ECS.

Step 3 Configure security group rules for the cluster.

1. On the **Dashboard** tab page, click **Add Security Group Rule**. In the **Add Security Group Rule** dialog box that is displayed, click **Manage Security Group Rule**.
2. Click the **Inbound Rules** tab, and click **Add Rule**. In the **Add Inbound Rule** dialog box, configure the IP address of the ECS and enable all ports.
3. After the security group rule is added, you can download and install the client on the ECS..
4. Use the client.

   Log in to the client node as the client installation user and run the following command to switch to the client directory:

   cd /opt/hadoopclient

   Run the following command to load environment variables:

   source bigdata_env

   If Kerberos authentication is enabled for the cluster, run the following command to authenticate the user. If Kerberos authentication is disabled for the current cluster, authentication is not required.

   kinit *MRS cluster user*

   Example:

   kinit admin

   Run the client command of a component.

Example:

Run the following command to view files in the HDFS root directory:

hdfs dfs -ls /

```
Found 15 items
drwxrwx--x   - hive       hive              0 2021-10-26 16:30 /apps
drwxr-xr-x   - hdfs       hadoop            0 2021-10-18 20:54 /datasets
drwxr-xr-x   - hdfs       hadoop            0 2021-10-18 20:54 /datastore
drwxrwx---+  - flink      hadoop            0 2021-10-18 21:10 /flink
drwxr-x---   - flume      hadoop            0 2021-10-18 20:54 /flume
drwxrwx--x   - hbase      hadoop            0 2021-10-30 07:31 /hbase
...
```

----End

14.11 Big Data Service Development

14.11.1 Can MRS Run Multiple Flume Tasks at a Time?

The Flume client supports multiple independent data flows. You can configure and link multiple sources, channels, and sinks in the **properties.properties** configuration file. These components can be linked to form multiple flows.

The following is an example of configuring two data flows in a configuration file:

```
server.sources = source1 source2
server.sinks = sink1 sink2
server.channels = channel1 channel2

#dataflow1
server.sources.source1.channels = channel1
server.sinks.sink1.channel = channel1

#dataflow2
server.sources.source2.channels = channel2
server.sinks.sink2.channel = channel2
```

14.11.2 How Do I Change FlumeClient Logs to Standard Logs?

1. Log in to the node where FlumeClient is running.
2. Go to the FlumeClient installation directory.

   For example, if the FlumeClient installation directory is **/opt/FlumeClient**, run the following command:

   cd /opt/FlumeClient/fusioninsight-flume-1.9.0/bin
3. Run the **./flume-manage.sh stop** command to stop FlumeClient.
4. Run the **vi /log4j.properties** command to open the **log4j.properties** file and change the value of **flume.root.logger** to **${flume.log.level},console**.
5. Run the **vim /flume-manager.sh** command to open the **flume-manager.sh** script in the **bin** directory in the Flume installation directory.
6. Comment out the following information in the **flume-manager.sh** script:

   >/dev/null 2>&1 &
7. Run the **./flume-manage.sh start** command to restart FlumeClient.

8. After the modification, check whether the Docker configuration is correct.

14.11.3 Where Are the .jar Files and Environment Variables of Hadoop Located?

- The **hadoopstreaming.jar** file is stored in the **/opt/share/hadoop-streaming-*** directory. ***** indicates the Hadoop version.
- The JDK environment variables are stored in **/opt/client/JDK/component_env**.
- The Hadoop environment variables are stored in **/opt/client/HDFS/component_env**.
- The Hadoop client path is **/opt/client/HDFS/hadoop**.

14.11.4 What Compression Algorithms Does HBase Support?

HBase supports the Snappy, LZ4, and gzip compression algorithms.

14.11.5 Can MRS Write Data to HBase Through the HBase External Table of Hive?

No. Hive on HBase supports only data query.

14.11.6 How Do I View HBase Logs?

1. Log in to the Master node in the cluster as user **root**.
2. Run the **su - omm** command to switch to user **omm**.
3. Run the **cd /var/log/Bigdata/hbase/** command to go to the **/var/log/Bigdata/hbase/** directory and view HBase logs.

14.11.7 How Do I Set the TTL for an HBase Table?

- Set the time to live (TTL) when creating a table:

  Create the **t_task_log** table, set the column family to **f**, and set the TTL to **86400** seconds.

  ```
  create 't_task_log',{NAME => 'f', TTL=>'86400'}
  ```
- Set the TTL for an existing table:

  ```
  disable "t_task_log" #Disable the table (services must be stopped).
  alter "t_task_log",NAME=>'data',TTL=>'86400' # Set the TTL value for the column family data.
  enable "t_task_log" #Restore the table.
  ```

14.11.8 How Do I Balance HDFS Data?

1. Log in to the master node of the cluster and run the corresponding command to configure environment variables. **/opt/client** indicates the client installation directory. Replace it with the actual one.

   source /opt/client/bigdata_env

   kinit Component service user (If Kerberos authentication is enabled for the cluster, run this command to authenticate the user. Skip this step if the Kerberos authentication is disabled.)
2. Run the following command to start the balancer:

/opt/client/HDFS/hadoop/sbin/start-balancer.sh -threshold 5

3. View the log.

   After you execute the balance task, the **hadoop-root-balancer-***Host name***.log** log file will be generated in the client installation directory **/opt/client/HDFS/hadoop/logs**.

4. (Optional) If you do not want to perform data balancing, run the following commands to stop the balancer:

   source /opt/client/bigdata_env

   kinit Component service user (If Kerberos authentication is enabled for the cluster, run this command to authenticate the user. Skip this step if the Kerberos authentication is disabled.)

   /opt/client/HDFS/hadoop/sbin/stop-balancer.sh -threshold 5

14.11.9 How Do I Change the Number of HDFS Replicas?

1. Go to the HDFS service configuration page.
   - For versions earlier than MRS 2.0.1, log in to MRS Manager, choose **Services** > **HDFS** > **Service Configuration**, and select **All** from the **Basic** drop-down list.
   - For MRS 2.0.1 or later, click the cluster name on the MRS console, choose **Components** > **HDFS** > **Service Configuration**, and select **All** from the **Basic** drop-down list.

     NOTE

     If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)
   - MRS 3.*x* or later: Log in to FusionInsight Manager. And choose **Cluster** > *Name of the desired cluster* > **Services** > **HDFS** > **Configurations** > **All Configurations**.
2. Search for **dfs.replication**, change the value (value range: 1 to 16), and restart the HDFS instance.

14.11.10 What Is the Port for Accessing HDFS Using Python?

The default port of open source HDFS is **50070** for versions earlier than MRS 3.0.0, and **9870** for MRS 3.0.0 or later. **Common HDFS Ports** describes the common ports of HDFS.

Common HDFS Ports

The protocol type of all ports in the table is TCP.

Parameter	Default Port	Port Description
dfs.namenode.rpc.port	- 9820 (versions earlier than MRS 3.*x*) - 8020 (MRS 3.*x* and later)	NameNode RPC port This port is used for: 1. Communication between the HDFS client and NameNode 2. Connection between the DataNode and NameNode **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
dfs.namenode.http.port	9870	HDFS HTTP port (NameNode) This port is used for: 1. Point-to-point NameNode checkpoint operations. 2. Connecting the remote web client to the NameNode UI **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
dfs.namenode.https.port	9871	HDFS HTTPS port (NameNode) This port is used for: 1. Point-to-point NameNode checkpoint operations 2. Connecting the remote web client to the NameNode UI **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Parameter	Default Port	Port Description
dfs.datanode.ipc.port	9867	IPC server port of DataNode This port is used for: Connection between the client and DataNode to perform RPC operations. **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
dfs.datanode.port	9866	DataNode data transmission port This port is used for: 1. Transmitting data from HDFS client from or to the DataNode 2. Point-to-point DataNode data transmission **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
dfs.datanode.http.port	9864	DataNode HTTP port This port is used for: Connecting to the DataNode from the remote web client in security mode **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Parameter	Default Port	Port Description
dfs.datanode.https.port	9865	HTTPS port of DataNode This port is used for: Connecting to the DataNode from the remote web client in security mode **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
dfs.JournalNode.rpc.port	8485	RPC port of JournalNode This port is used for: Client communication to access multiple types of information **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
dfs.journalnode.http.port	8480	JournalNode HTTP port This port is used for: Connecting to the JournalNode from the remote web client in security mode **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

Parameter	Default Port	Port Description
dfs.journalnode.https.port	8481	HTTPS port of JournalNode This port is used for: Connecting to the JournalNode from the remote web client in security mode **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes
httpfs.http.port	14000	Listening port of the HttpFS HTTP server This port is used for: Connecting to the HttpFS from the remote REST API **NOTE** The port ID is a recommended value and is specified based on the product. The port range is not restricted in the code. ● Is the port enabled by default during the installation: Yes ● Is the port enabled after security hardening: Yes

14.11.11 How Do I Modify the HDFS Active/Standby Switchover Class?

If the **org.apache.hadoop.hdfs.server.namenode.ha.AdaptiveFailoverProxyProvider** class is unavailable when a cluster of MRS 3.*x* connects to NameNodes using HDFS, the cause is that the HDFS active/standby switchover class of the cluster is configured improperly. To solve the problem, perform the following operations:

- Method 1: Add the **hadoop-plugins-*xxx*.jar** package to the **classpath** or **lib** directory of your program.

  The **hadoop-plugins-*xxx*.jar** package is stored in the HDFS client directory, for example, **$HADOOP_HOME/share/hadoop/common/lib/hadoop-plugins-8.0.2-302023.jar**.

- Method 2: Change the configuration item of HDFS to the corresponding open source class, as shown in the follows:

  dfs.client.failover.proxy.provider.hacluster=org.apache.hadoop.hdfs.server.namenode.ha.ConfiguredFailoverProxyProvider

14.11.12 What Is the Recommended Number Type of DynamoDB in Hive Tables?

smallint is recommended.

14.11.13 Can the Hive Driver Be Interconnected with DBCP2?

The Hive driver cannot be interconnected with the DBCP2 database connection pool. The DBCP2 database connection pool invokes the **isValid** method to check whether a connection is available. However, Hive directly throws an exception when implementing this method.

14.11.14 How Do I View the Hive Table Created by Another User?

Versions earlier than MRS 3.*x*:

1. Log in to MRS Manager and choose **System** > **Permission** > **Manage Role**.
2. Click **Create Role**, and set **Role Name** and **Description**.
3. In the **Permission** table, choose **Hive** > **Hive Read Write Privileges**.
4. In the database list, click the name of the database where the table created by user B is stored. The table is displayed.
5. In the **Permission** column of the table created by user B, select **SELECT**.
6. Click **OK**, and return to the **Role** page.
7. Choose **System** > **Manage User**. Locate the row containing user A, click **Modify** to bind the new role to user A, and click **OK**. After about 5 minutes, user A can access the table created by user B.

MRS 3.*x* or later:

1. Log in to FusionInsight Manager and choose **Cluster** > **Services**. On the page that is displayed, choose **Hive**. On the displayed page, choose **More**, and check whether **Enable Ranger** is grayed out.
   - If yes, go to **9**.
   - If no, perform **2** to **8**.
2. Log in to FusionInsight Manager and choose **System** > **Permission** > **Role**.
3. Click **Create Role**, and set **Role Name** and **Description**.
4. In the **Configure Resource Permission** table, choose *Name of the desired cluster* > **Hive** > **Hive Read Write Privileges**.
5. In the database list, click the name of the database where the table created by user B is stored. The table is displayed.
6. In the **Permission** column of the table created by user B, select **Select**.
7. Click **OK**, and return to the **Role** page.
8. Choose **Permission** > **User**. On the **Local User** page that is displayed, locate the row containing user A, click **Modify** in the **Operation** column to bind the new role to user A, and click **OK**. After about 5 minutes, user A can access the table created by user B.
9. Perform the following steps to add the Ranger access permission policy of Hive:

a. Log in to FusionInsight Manager as a Hive administrator and choose **Cluster** > **Services**. On the page that is displayed, choose **Ranger**. On the displayed page, click the URL next to **Ranger WebUI** to go to the Ranger management page.

b. On the home page, click the component plug-in name in the **HADOOP SQL** area, for example, **Hive**.

c. On the **Access** tab page, click **Add New Policy** to add a Hive permission control policy.

d. In the **Create Policy** dialog box that is displayed, set the following parameters:

   - **Policy Name**: Enter a policy name, for example, **table_test_hive**.
   - **database**: Enter or select the database where the table created by user B is stored, for example, **default**.
   - **table**: Enter or select the table created by user B, for example, **test**.
   - **column**: Enter and select a column, for example, *****.
   - In the **Allow Conditions** area, click **Select User**, select user A, click **Add Permissions**, and select **select**.
   - Click **Add**.

10. Perform the following steps to add the Ranger access permission policy of HDFS:

a. Log in to FusionInsight Manager as user **rangeradmin** and choose **Cluster** > **Services**. On the page that is displayed, choose **Ranger**. On the displayed page, click the URL next to **Ranger WebUI** to go to the Ranger management page.

b. On the home page, click the component plug-in name in the **HDFS** area, for example, **hacluster**.

c. Click **Add New Policy** to add a HDFS permission control policy.

d. In the **Create Policy** dialog box that is displayed, set the following parameters:

   - **Policy Name**: Enter a policy name, for example, **tablehdfs_test**.
   - **Resource Path**: Set this parameter to the HDFS path where the table created by user B is stored, for example, **/user/hive/warehouse/***Database name***/***Table name*.
   - In the **Allow Conditions** area, select user A for **Select User**, click **Add Permissions** in the **Permissions** column, and select **Read** and **Execute**.
   - Click **Add**.

11. View basic information about the policy in the policy list. After the policy takes effect, user A can view the table created by user B.

14.11.15 Can I Export the Query Result of Hive Data?

Run the following statement to export the query result of Hive data:

```
insert overwrite local directory "/tmp/out/" row format delimited fields terminated by "\t" select * from
table;
```

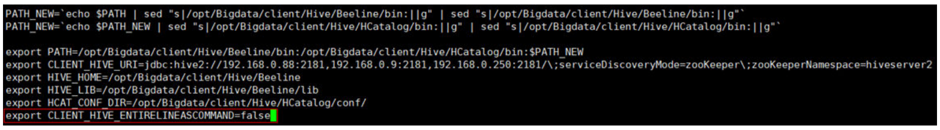
14.11.16 How Do I Do If an Error Occurs When Hive Runs the beeline -e Command to Execute Multiple Statements?

When Hive of MRS 3.*x* runs the **beeline -e " use default;show tables;"** command, the following error message is displayed: Error while compiling statement: FAILED: ParseException line 1:11 missing EOF at ';' near 'default' (state=42000,code=40000).

Solutions:

- Method 1: Replace the **beeline -e " use default;show tables;"** command with **beeline --entirelineascommand=false -e "use default;show tables;"**.
- Method 2:
  a. In the **/opt/Bigdata/client/Hive** directory on the Hive client, change **export CLIENT_HIVE_ENTIRELINEASCOMMAND=true** in the **component_env** file to **export CLIENT_HIVE_ENTIRELINEASCOMMAND=false**.

     Figure 14-1 Changing the **component_env** file

```
PATH_NEW=`echo $PATH | sed "s|/opt/Bigdata/client/Hive/Beeline/bin:||g" | sed "s|/opt/Bigdata/client/Hive/Beeline/bin:||g"`
PATH_NEW=`echo $PATH_NEW | sed "s|/opt/Bigdata/client/Hive/HCatalog/bin:||g" | sed "s|/opt/Bigdata/client/Hive/HCatalog/bin:||g"`

export PATH=/opt/Bigdata/client/Hive/Beeline/bin:/opt/Bigdata/client/Hive/HCatalog/bin:$PATH_NEW
export CLIENT_HIVE_URI=jdbc:hive2://192.168.0.88:2181,192.168.0.9:2181,192.168.0.250:2181/\;serviceDiscoveryMode=zooKeeper\;zooKeeperNamespace=hiveserver2
export HIVE_HOME=/opt/Bigdata/client/Hive/Beeline
export HIVE_LIB=/opt/Bigdata/client/Hive/Beeline/lib
export HCAT_CONF_DIR=/opt/Bigdata/client/Hive/HCatalog/conf/
export CLIENT_HIVE_ENTIRELINEASCOMMAND=false
```

  b. Run the following command to verify the configuration:

     source /opt/Bigdata/client/bigdata_env

     beeline -e " use default;show tables;"

14.11.17 How Do I Do If a "hivesql/hivescript" Job Fails to Submit After Hive Is Added?

This issue occurs because the **MRS CommonOperations** permission bound to the user group to which the user who submits the job belongs does not include the Hive permission after being synchronized to Manager. To solve this issue, perform the following operations:

1. Add the Hive service.
2. Log in to the IAM console and create a user group. The policy bound to the user group is the same as that of the user group to which the user who submits the job belongs.
3. Add the user who submits the job to the new user group.
4. Refresh the cluster details page on the MRS console. The status of IAM user synchronization is **Not synchronized**.
5. Click **Synchronize** on the right of **IAM User Sync**. Go back to the previous page. In the navigation pane on the left, choose **Operation Logs** and check whether the user is changed.
   - If yes, submit the Hive job again.

- If no, check whether all the preceding operations are complete.
  - If yes, contact the O&M personnel.
  - If no, submit the Hive job after the preceding operations are complete.

14.11.18 What If an Excel File Downloaded on Hue Failed to Open?

1. Log in to a Master node as user **root** and switch to user **omm**.

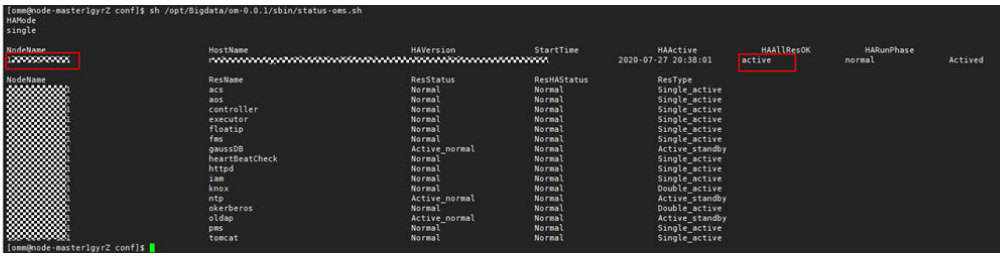
   su - omm

2. Check whether the current node is the active OMS node.

   sh ${BIGDATA_HOME}/om-0.0.1/sbin/status-oms.sh

   If **active** is displayed in the command output, the node is the active node. Otherwise, log in to the other Master node.

   Figure 14-2 Active OMS node



```
[omm@node-master1gyrZ conf]$ sh /opt/Bigdata/om-0.0.1/sbin/status-oms.sh
HAMode
single

NodeName              HostName              HAVersion        StartTime             HAActive     HAAllResOK     HARunPhase
[redacted]            [redacted]                             2020-07-27 20:38:01   active       normal         Actived

NodeName              ResName               ResStatus        ResHAStatus           ResType
[redacted]            acs                   Normal           Normal                Single_active
                      aos                   Normal           Normal                Single_active
                      controller            Normal           Normal                Single_active
                      executor              Normal           Normal                Single_active
                      floatip               Normal           Normal                Single_active
                      fms                   Normal           Normal                Single_active
                      gaussDB               Active_normal    Normal                Active_standby
                      heartBeatCheck        Normal           Normal                Single_active
                      httpd                 Normal           Normal                Single_active
                      iam                   Normal           Normal                Single_active
                      knox                  Normal           Normal                Double_active
                      ntp                   Active_normal    Normal                Active_standby
                      okerberos             Normal           Normal                Double_active
                      oldap                 Active_normal    Normal                Active_standby
                      pms                   Normal           Normal                Single_active
                      tomcat                Normal           Normal                Single_active
[omm@node-master1gyrZ conf]$
```

3. Go to the **${BIGDATA_HOME}/Apache-httpd-*/conf** directory.

   cd ${BIGDATA_HOME}/Apache-httpd-*/conf

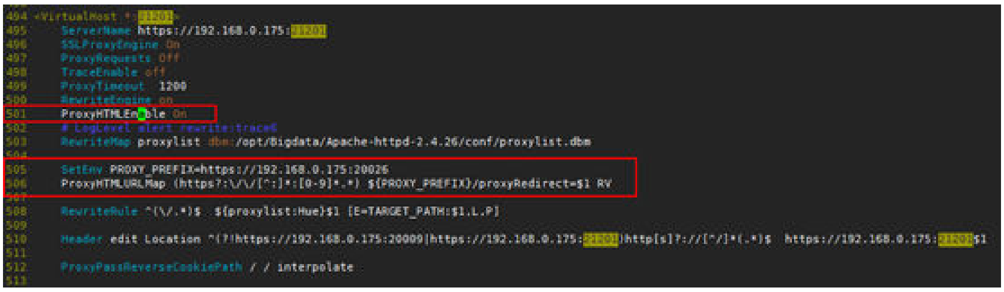
4. Open the **httpd.conf** file.

   vim httpd.conf

5. Search for **21201** in the file and delete the following content from the file (The values of *proxy_ip* and *proxy_port* in **Figure 14-3** are examples only):

```
ProxyHTMLEnable On
SetEnv PROXY_PREFIX=https://[proxy_ip]:[proxy_port]
ProxyHTMLURLMap (https?:\/\/[^:]*:[0-9]*.*) ${PROXY_PREFIX}/proxyRedirect=$1 RV
```

   Figure 14-3 Content to be deleted



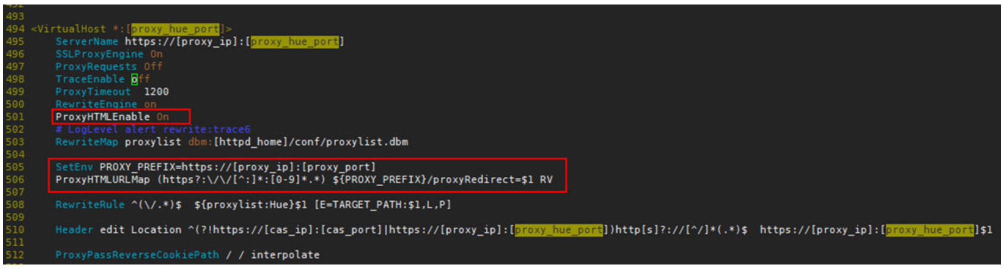
```
<VirtualHost *:21201>
    ServerName https://192.168.0.175:21201
    SSLProxyEngine On
    ProxyRequests Off
    TraceEnable off
    ProxyTimeout  1200
    RewriteEngine on
    ProxyHTMLEnable On
    # LogLevel alert rewrite:trace6
    RewriteMap proxylist dbm:/opt/Bigdata/Apache-httpd-2.4.26/conf/proxylist.dbm

    SetEnv PROXY_PREFIX=https://192.168.0.175:20026
    ProxyHTMLURLMap (https?:\/\/[^:]*:[0-9]*.*) ${PROXY_PREFIX}/proxyRedirect=$1 RV

    RewriteRule ^(\/.*)$  ${proxylist:Hue}$1 [E=TARGET_PATH:$1,L,P]

    Header edit Location ^(?!https://192.168.0.175:20009|https://192.168.0.175:21201)http[s]?://[^/]*(.*)$  https://192.168.0.175:21201$1

    ProxyPassReverseCookiePath / / interpolate
```

6. Save the modification and exit.

7. Open the **httpd.conf** file again, search for **proxy_hue_port**, and delete the following content:

```
ProxyHTMLEnable On
SetEnv PROXY_PREFIX=https://[proxy_ip]:[proxy_port]
ProxyHTMLURLMap (https?:\/\/[^:]*:[0-9]*.*) ${PROXY_PREFIX}/proxyRedirect=$1 RV
```

Figure 14-4 Content to be deleted



```
<VirtualHost *:[proxy_hue_port]>
    ServerName https://[proxy_ip]:[proxy_hue_port]
    SSLProxyEngine On
    ProxyRequests Off
    TraceEnable Off
    ProxyTimeout  1200
    RewriteEngine on
    ProxyHTMLEnable On
    # LogLevel alert rewrite:trace6
    RewriteMap proxylist dbm:[httpd_home]/conf/proxylist.dbm

    SetEnv PROXY_PREFIX=https://[proxy_ip]:[proxy_port]
    ProxyHTMLURLMap (https?:\/\/[^:]*:[0-9]*.*) ${PROXY_PREFIX}/proxyRedirect=$1 RV

    RewriteRule ^(\/.*)$  ${proxylist:Hue}$1 [E=TARGET_PATH:$1,L,P]

    Header edit Location ^(?!https://[cas_ip]:[cas_port]|https://[proxy_ip]:[proxy_hue_port])http[s]?://[^/]*(.*)$  https://[proxy_ip]:[proxy_hue_port]$1

    ProxyPassReverseCookiePath / / interpolate
```

8. Save the modification and exit.
9. Run the following command to restart the **httpd** process:

   sh ${BIGDATA_HOME}/Apache-httpd-*/setup/restarthttpd.sh
10. Check whether the **httpd.conf** file on the standby Master node is modified. If the file is modified, no further action is required. If the file is not modified, modify the **httpd.conf** file on the standby Master node in the same way. You do not need to restart the **httpd** process.
11. Download the Excel file again. You can open the file successfully.

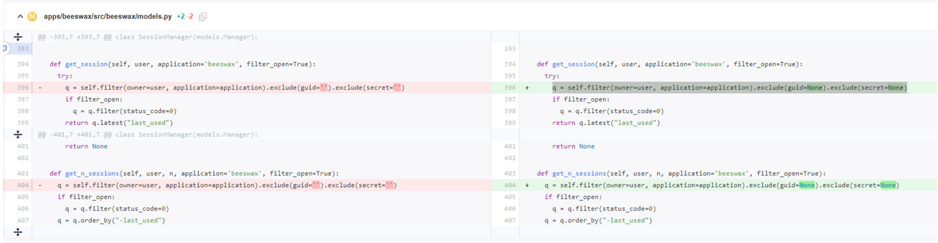
14.11.19 How Do I Do If Sessions Are Not Released After Hue Connects to HiveServer and the Error Message "over max user connections" Is Displayed?

Applicable versions: MRS 3.1.0 and earlier

1. Modify the following file on the two Hue nodes:

   /opt/Bigdata/FusionInsight_Porter_8.*/install/FusionInsight-Hue-*/hue/apps/beeswax/src/beeswax/models.py
2. Change the configurations in lines 396 and 404.

   Change **q Changed = self.filter(owner=user, application=application).exclude(guid='').exclude(secret='')** to **q = self.filter(owner=user, application=application).exclude(guid=None).exclude(secret=None)**.



apps/beeswax/src/beeswax/models.py +2 -2

```
@@ -393,7 +393,7 @@ class SessionManager(models.Manager):
393
394       def get_session(self, user, application='beeswax', filter_open=True):
395         try:
396  -        q = self.filter(owner=user, application=application).exclude(guid='').exclude(secret='')
397           if filter_open:
398             q = q.filter(status_code=0)
399           return q.latest("last_used")
@@ -401,7 +401,7 @@ class SessionManager(models.Manager):
401           return None
402
403       def get_n_sessions(self, user, n, application='beeswax', filter_open=True):
404  -      q = self.filter(owner=user, application=application).exclude(guid='').exclude(secret='')
405         if filter_open:
406           q = q.filter(status_code=0)
407         q = q.order_by("-last_used")
```

```
393
394       def get_session(self, user, application='beeswax', filter_open=True):
395         try:
396  +        q = self.filter(owner=user, application=application).exclude(guid=None).exclude(secret=None)
397           if filter_open:
398             q = q.filter(status_code=0)
399           return q.latest("last_used")

401           return None
402
403       def get_n_sessions(self, user, n, application='beeswax', filter_open=True):
404  +      q = self.filter(owner=user, application=application).exclude(guid=None).exclude(secret=None)
405         if filter_open:
406           q = q.filter(status_code=0)
407         q = q.order_by("-last_used")
```

14.11.20 How Do I Reset Kafka Data?

You can reset Kafka data by deleting Kafka topics.

- Delete a topic: **kafka-topics.sh --delete --zookeeper** *ZooKeeper Cluster service IP address*:**2181/kafka --topic *topicname***

- Query all topics: **kafka-topics.sh --zookeeper** *ZooKeeper cluster service IP address*:**2181/kafka --list**

After the deletion command is executed, empty topics will be deleted immediately. If a topic has data, the topic will be marked for deletion and will be deleted by Kafka later.

14.11.21 How Do I Obtain the Client Version of MRS Kafka?

Run the **--bootstrap-server** command to query the information about the client.

14.11.22 What Access Protocols Are Supported by Kafka?

Kafka supports PLAINTEXT, SSL, SASL_PLAINTEXT, and SASL_SSL.

14.11.23 How Do I Do If Error Message "Not Authorized to access group xxx" Is Displayed When a Kafka Topic Is Consumed?

This issue is caused by the conflict between the Ranger authentication and ACL authentication of a cluster. If a Kafka cluster uses ACL for permission access control and Ranger authentication is enabled for the Kafka component, all authentications of the component are managed by Ranger. The permissions set by the original authentication plug-in are invalid. As a result, ACL authorization does not take effect. You can disable Ranger authentication of Kafka and restart the Kafka service to rectify the fault. The procedure is as follows:

1. Log in to FusionInsight Manager and choose **Cluster** > **Services** > **Kafka**.
2. In the upper right corner of the **Dashboard** page, click **More** and choose **Disable Ranger**. In the displayed dialog box, enter the password and click **OK**. After the operation is successful, click **Finish**.
3. In the upper right corner of the **Dashboard** page, click **More** and choose **Restart Service** to restart the Kafka service.

14.11.24 What Compression Algorithms Does Kudu Support?

Kudu supports **Snappy**, **LZ4**, and **zlib**. **LZ4** is used by default.

14.11.25 How Do I View Kudu Logs?

1. Log in to the Master node in the cluster.
2. Run the **su - omm** command to switch to user **omm**.
3. Run the **cd /var/log/Bigdata/kudu/** command to go to the **/var/log/Bigdata/kudu/** directory and view Kudu logs.

14.11.26 How Do I Handle the Kudu Service Exceptions Generated During Cluster Creation?

Viewing the Kudu Service Exception Logs

1. Log in to the MRS console.

2. Click the name of the cluster.
3. On the page displayed, choose **Components** > **Kudu** > **Instances** and locate the IP address of the abnormal instance.

   If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)
4. Log in to the node where the abnormal instance resides, and view the Kudu log.
   ```
   cd /var/log/Bigdata/Kudu
   [root@node-master1AERu kudu]# ls
   healthchecklog  runninglog  startlog
   ```
   You can find the Kudu health check logs in the **healthchecklog** directory, the startup logs in the **startlog** directory, and the Kudu process run logs in the **runninglog** directory.
   ```
   [root@node-master1AERu logs]# pwd
   /var/log/Bigdata/kudu/runninglog/master/logs
   [root@node-master1AERu logs]# ls -al
   kudu-master.ERROR   kudu-master.INFO   kudu-master.WARNING
   ```
   Run logs are classified into three types: ERROR, INFO, and WARNING. Each type of run logs is recorded in the corresponding file. You can run the **cat** command to view run logs of each type.

Handling Kudu Service Exceptions

The **/var/log/Bigdata/kudu/runninglog/master/logs/kudu-master.INFO** file contains the following error information:

```
"Unable to init master catalog manager: not found: Unable to initialize catalog manager: Failed to initialize sys tables async: Unable to load consensus metadata for tablet 0000000000000000000000: xxx"
```

If this exception occurs when the Kudu service is installed for the first time, the KuduMaster service is not started. The data inconsistency causes the startup failure. To solve the problem, perform the following steps to clear the data directories and restart the Kudu service. If the Kudu service is not installed for the first time, clearing the data directories will cause data loss. In this case, migrate data and clear the data directory.

1. Search for the data directories **fs_data_dir**, **fs_wal_dir**, and **fs_meta_dir**.

   find /opt -name master.gflagfile

   cat /opt/Bigdata/FusionInsight_Kudu_*/*_KuduMaster/etc/master.gflagfile | grep fs_
2. On the cluster details page, choose **Components** > **Kudu** and click **Stop Service**.
3. Clear the Kudu data directories on all KuduMaster and KuduTserver nodes. The following command uses two data disks as an example.

   rm -Rvf /srv/Bigdata/data1/kudu, rm -Rvf /srv/Bigdata/data2/kudu
4. On the cluster details page, choose **Components** > **Kudu** and choose **More** > **Restart Service**.
5. Check the Kudu service status and logs.

14.11.27 Does OpenTSDB Support Python APIs?

OpenTSDB supports Python APIs. OpenTSDB provides HTTP-based RESTful APIs that are language-independent. Any language that supports HTTP requests can interconnect to OpenTSDB.

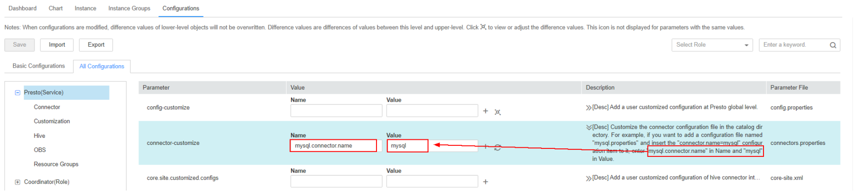
14.11.28 How Do I Configure Other Data Sources on Presto?

In this section, MySQL is used as an example.

- For MRS 1.*x* and 3.*x* clusters, do the following:
  a. Log in to the MRS management console.
  b. Click the name of the cluster to go to its details page.
  c. Click the **Components** tab and then **Presto** in the component list. On the page that is displayed, click the **Configurations** tab then the **All Configurations** sub-tab.
  d. On the Presto configuration page that is displayed, find **connector-customize**.
  e. Set **Name** and **Value** as follows:

     Name: **mysql.connector.name**

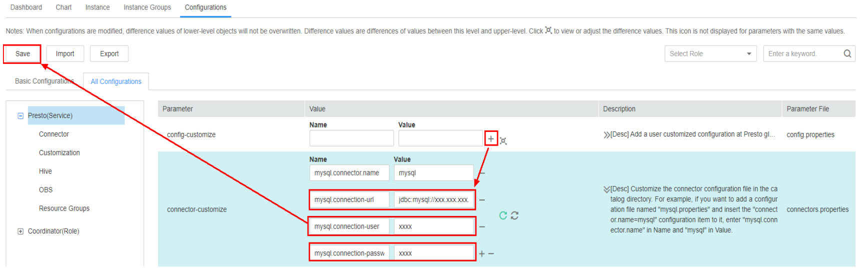
     Value: **mysql**



Notes: When configurations are modified, difference values of lower-level objects will not be overwritten. Difference values are differences of values between this level and upper-level. Click ⌘ to view or adjust the difference values. This icon is not displayed for parameters with the same values.

Parameter	Value	Description	Parameter File
config-customize	Name: Value:	[Desc] Add a user customized configuration at Presto global level.	config.properties
connector-customize	Name: mysql.connector.name Value: mysql	[Desc] Customize the connector configuration file in the catalog directory. For example, if you want to add a configuration file named "mysql.properties" and insert the "connector.name=mysql" configuration item to it, enter "mysql.connector.name" in Name and "mysql" in Value.	connectors.properties
core.site.customized.configs	Name: Value:	[Desc] Add a user customized configuration of hive connector int...	core-site.xml

  f. Click the plus sign (+) to add three more fields and set **Name** and **Value** according to the table below. Then click **Save**.



Notes: When configurations are modified, difference values of lower-level objects will not be overwritten. Difference values are differences of values between this level and upper-level. Click ⌘ to view or adjust the difference values. This icon is not displayed for parameters with the same values.

Parameter	Value	Description	Parameter File
config-customize	Name: Value:	[Desc] Add a user customized configuration at Presto gl...	config.properties
connector-customize	mysql.connector.name = mysql mysql.connection-url = jdbc:mysql://xxx.xxx.xxx. mysql.connection-user = xxxx mysql.connection-passw = xxxx	[Desc] Customize the connector configuration file in the catalog directory. For example, if you want to add a configuration file named "mysql.properties" and insert the "connector.name=mysql" configuration item to it, enter "mysql.connector.name" in Name and "mysql" in Value.	connectors.properties

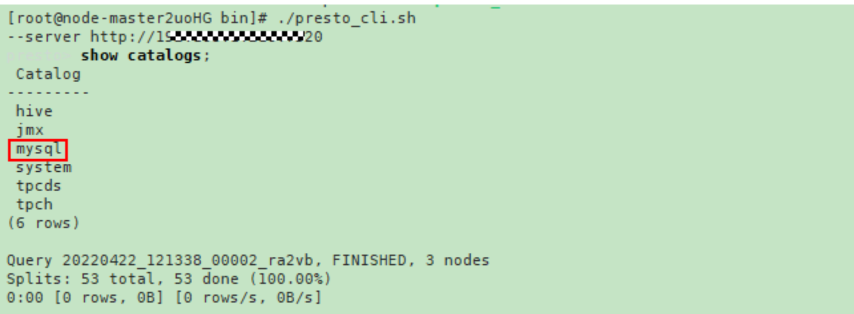
Name	Value	Description
mysql.connection-url	jdbc:mysql://xxx.xxx.xxx.xxx:3306	Database connection pool
mysql.connection-user	xxxx	Database username
mysql.connection-password	xxxx	Database password

  g. Restart the Presto service.

h. Run the following command to connect to the Presto Server of the cluster:

presto_cli.sh --*krb5-config-path* {krb5.conf path} --*krb5-principal* {User principal} --*krb5-keytab-path* {user.keytab path} --*user* {presto username}

i. Log in to Presto and run the **show catalogs** command to check whether the data source list mysql of Presto can be queried.



```
[root@node-master2uoHG bin]# ./presto_cli.sh
--server http://1[redacted]20
presto> show catalogs;
 Catalog
---------
 hive
 jmx
 mysql
 system
 tpcds
 tpch
(6 rows)

Query 20220422_121338_00002_ra2vb, FINISHED, 3 nodes
Splits: 53 total, 53 done (100.00%)
0:00 [0 rows, 0B] [0 rows/s, 0B/s]
```

Run the **show schemas from mysql** command to query the MySQL database.

- For MRS 2.*x* clusters, do the following:

  a. Create the **mysql.properties** configuration file containing the following content:

  connector.name=mysql

  connection-url=jdbc:mysql://mysqlIp:3306

  connection-user=Username

  connection-password=Password

  NOTE

  - **mysqlIp** indicates the IP address of the MySQL instance, which must be able to communicate with the MRS network.
  - The username and password are those used to log in to the MySQL database.

  b. Upload the configuration file to the **/opt/Bigdata/MRS_Current/1_14_Coordinator/etc/catalog/** directory on the master node (where the Coordinator instance resides) and the **/opt/Bigdata/MRS_Current/1_14_Worker/etc/catalog/** directory on the core node (depending on the actual directory in the cluster), and change the file owner group to **omm:wheel**.

  c. Restart the Presto service.

14.11.29 How Do I Connect to Spark Shell from MRS?

1. Log in to the Master node in the cluster as user **root**.
2. Run the following command to configure environment variables:

   source /opt/client/bigdata_env

3. If Kerberos authentication is enabled for the cluster, authenticate the user. If Kerberos authentication is disabled, skip this step.

   Command: **kinit** *MRS cluster user*

   Example:

- If the user is a machine-machine user, run **kinit -kt user.keytab sparkuser**.
- If the user is a human-machine user, run **kinit sparkuser**.

4. Run the following command to connect to Spark shell:

   spark-shell

14.11.30 How Do I Connect to Spark Beeline from MRS?

1. Log in to the master node in the cluster as user **root**.
2. Run the following command to configure environment variables:

   source /opt/client/bigdata_env

3. If Kerberos authentication is enabled for the cluster, authenticate the user. If Kerberos authentication is disabled, skip this step.

   Command: **kinit** *MRS cluster user*

   Example:

   - If the user is a machine-machine user, run **kinit -kt user.keytab sparkuser**.
   - If the user is a human-machine user, run **kinit sparkuser**.

4. Run the following command to connect to Spark Beeline:

   spark-beeline

5. Run commands on Spark Beeline. For example, create the table **test** in the **obs://mrs-word001/table/** directory.

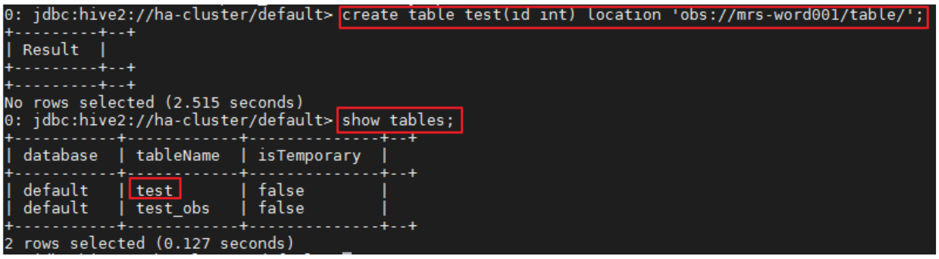
   create table test(id int) location 'obs://mrs-word001/table/';

6. Query all tables.

   show tables;

   If the table **test** is displayed in the command output, OBS is successfully accessed.

   Figure 14-5 Returned table name



```
0: jdbc:hive2://ha-cluster/default> create table test(id int) location 'obs://mrs-word001/table/';
+---------+--+
| Result  |
+---------+--+
+---------+--+
No rows selected (2.515 seconds)
0: jdbc:hive2://ha-cluster/default> show tables;
+-----------+------------+--------------+--+
| database  | tableName  | isTemporary  |
+-----------+------------+--------------+--+
| default   | test       | false        |
| default   | test_obs   | false        |
+-----------+------------+--------------+--+
2 rows selected (0.127 seconds)
```

7. Press **Ctrl+C** to exit the Spark Beeline.

14.11.31 Where Are the Execution Logs of Spark Jobs Stored?

- Logs of unfinished Spark jobs are stored in the **/srv/BigData/hadoop/data1/nm/containerlogs/** directory on the Core node.
- Logs of finished Spark jobs are stored in the **/tmp/logs/***username***/logs** directory of HDFS.

14.11.32 How Do I Specify a Log Path When Submitting a Task in an MRS Storm Cluster?

You can modify the **/opt/Bigdata/MRS_*XXX* /1_*XX* _Supervisor/etc/worker.xml** file on the streaming Core node of MRS, set the value of **filename** to the path, and restart the corresponding instance on Manager.

You are advised not to modify the default log configuration of MRS. Otherwise, the log system may become abnormal.

14.11.33 How Do I Check Whether the ResourceManager Configuration of Yarn Is Correct?

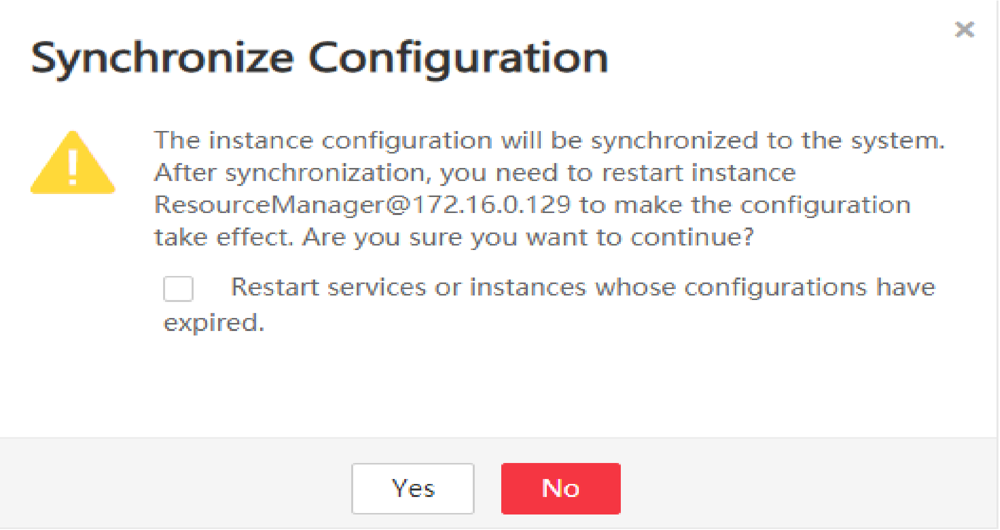
Step 1 Log in to MRS Manager and choose **Services** > **Yarn** > **Instance**.

Step 2 Synchronize the configuration between the two ResourceManager nodes.

Perform the following steps on each ResourceManager node:

1. Click the name of the ResourceManager node, and choose **More** > **Synchronize Configuration**.
2. In the dialog box displayed, deselect **Restart services or instances whose configurations have expired** and click **Yes**.

Figure 14-6 Synchronization configurations



Synchronize Configuration

The instance configuration will be synchronized to the system. After synchronization, you need to restart instance ResourceManager@172.16.0.129 to make the configuration take effect. Are you sure you want to continue?

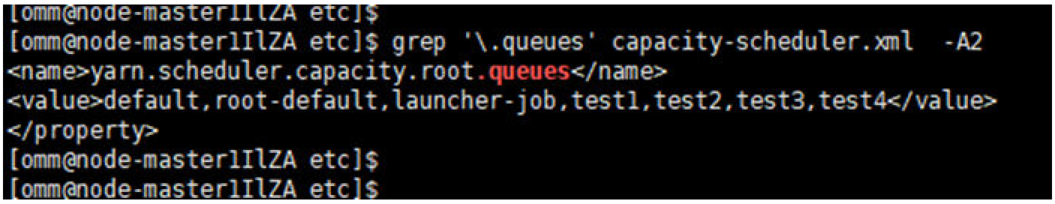
□ Restart services or instances whose configurations have expired.

Step 3 Log in to the Master nodes as user **root**.

Step 4 Run the **cd /opt/Bigdata/MRS_Current/*_*_ResourceManager/etc_UPDATED/** command to go to the **etc_UPDATED** directory.

Step 5 Run the **grep '\.queues' capacity-scheduler.xml -A2** command to display all configured queues and check whether the queues are consistent with those displayed on Manager.

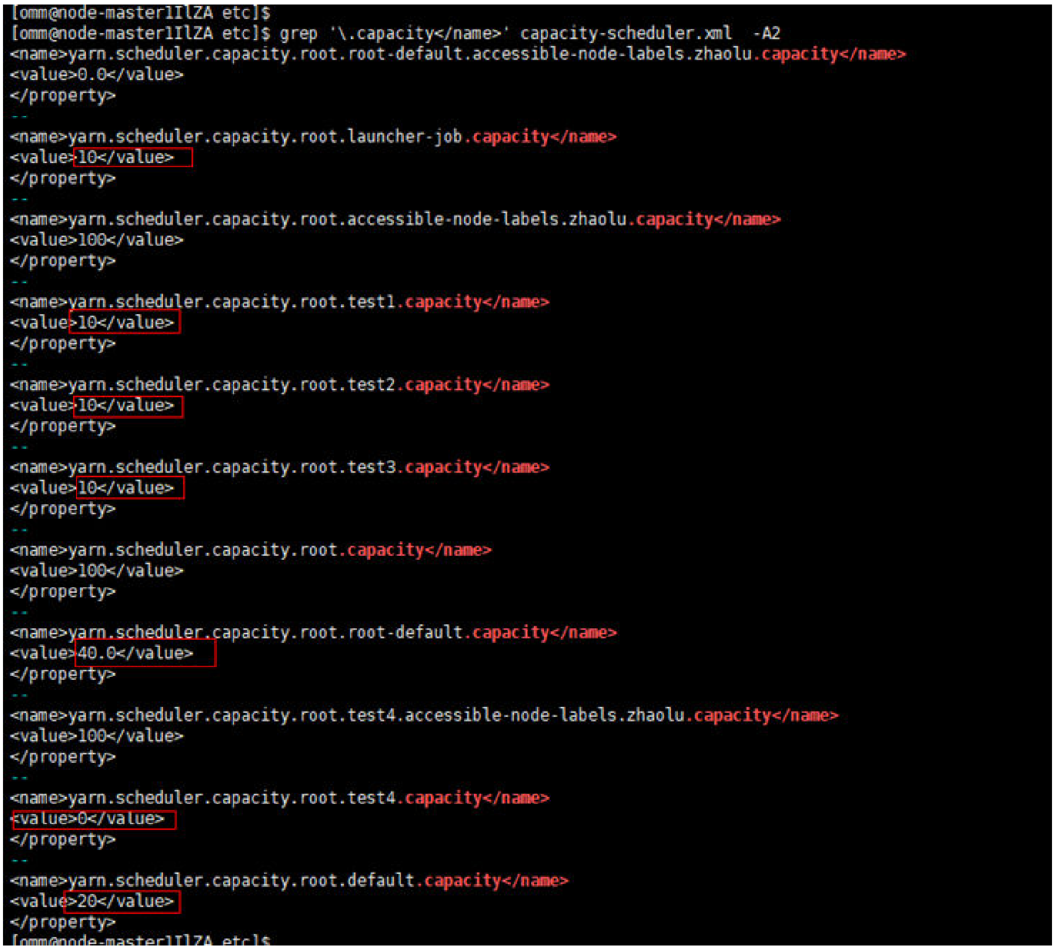
root-default is hidden on the Manager page.



```
[omm@node-master1IlZA etc]$
[omm@node-master1IlZA etc]$ grep '\.queues' capacity-scheduler.xml  -A2
<name>yarn.scheduler.capacity.root.queues</name>
<value>default,root-default,launcher-job,test1,test2,test3,test4</value>
</property>
[omm@node-master1IlZA etc]$
[omm@node-master1IlZA etc]$
```

Step 6 Run the **grep '\.capacity</name>' capacity-scheduler.xml -A2** command to display the value of each queue and check whether the value of each queue is the same as that displayed on Manager. Check whether the sum of the values configured for all queues is **100**.

- If the sum is **100**, the configuration is correct.
- If the sum is not **100**, the configuration is incorrect. Perform the following steps to rectify the fault.



```
[omm@node-master1IlZA etc]$
[omm@node-master1IlZA etc]$ grep '\.capacity</name>' capacity-scheduler.xml  -A2
<name>yarn.scheduler.capacity.root.root-default.accessible-node-labels.zhaolu.capacity</name>
<value>0.0</value>
</property>
--
<name>yarn.scheduler.capacity.root.launcher-job.capacity</name>
<value>10</value>
</property>
--
<name>yarn.scheduler.capacity.root.accessible-node-labels.zhaolu.capacity</name>
<value>100</value>
</property>
--
<name>yarn.scheduler.capacity.root.test1.capacity</name>
<value>10</value>
</property>
--
<name>yarn.scheduler.capacity.root.test2.capacity</name>
<value>10</value>
</property>
--
<name>yarn.scheduler.capacity.root.test3.capacity</name>
<value>10</value>
</property>
--
<name>yarn.scheduler.capacity.root.capacity</name>
<value>100</value>
</property>
--
<name>yarn.scheduler.capacity.root.root-default.capacity</name>
<value>40.0</value>
</property>
--
<name>yarn.scheduler.capacity.root.test4.accessible-node-labels.zhaolu.capacity</name>
<value>100</value>
</property>
--
<name>yarn.scheduler.capacity.root.test4.capacity</name>
<value>0</value>
</property>
--
<name>yarn.scheduler.capacity.root.default.capacity</name>
<value>20</value>
</property>
[omm@node-master1IlZA etc]$
```

Step 7 Log in to MRS Manager, and select **Hosts**.

Step 8 Determine the active Master node. The host name of the active Master node starts with a solid pentagon.

Step 9 Log in to the active Master node as user **root**.

Step 10 Run the **su - omm** command to switch to user **omm**.

Step 11 Run the **sh /opt/Bigdata/om-0.0.1/sbin/restart-controller.sh** command to restart the controller when no operation is being performed on Manager.

Restarting the controller will not affect the big data component services.

Step 12 Repeat **Step 1** to **Step 6** to synchronize ResourceManager configurations and check whether the configurations are correct.

If the latest configuration has not been loaded after the configuration synchronization is complete, a message will be displayed on the Manager page indicating that the configuration has expired. However, this will not affect services. The latest configuration will be automatically loaded when the component restarts.

----End

14.11.34 How Do I Modify the allow_drop_detached Parameter of ClickHouse?

Step 1 Log in to the node where the ClickHouse client is located as user **root**.

Step 2 Run the following commands to go to the client installation directory and set the environment variables:

cd /opt/*Client installation directory*

source bigdata_env

Step 3 If Kerberos authentication is enabled for the cluster, run the following command to authenticate the user. If Kerberos authentication is disabled, skip this step.

kinit *MRS cluster user*

NOTE

The user must have the ClickHouse administrator permissions.

Step 4 Run the **clickhouse client --host** *192.168.42.90* **--secure -m** command, in which *192.168.42.90* indicates the IP address of the ClickHouseServer instance node. The command output is as follows:

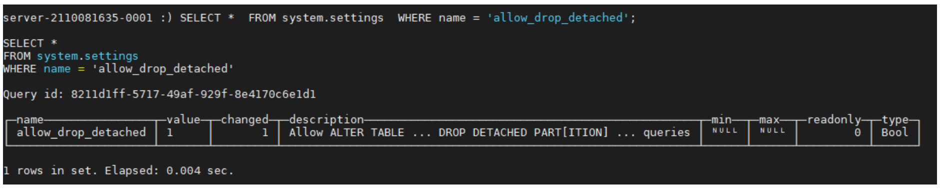
```
[root@server-2110082001-0017 hadoopclient]# clickhouse client --host 192.168.42.90 --secure -m
ClickHouse client version 21.3.4.25.
Connecting to 192.168.42.90:21427.
Connected to ClickHouse server version 21.3.4 revision 54447.
```

Step 5 Run the following command to set the value of the **allow_drop_detached** parameter, for example, **1**:

set allow_drop_detached=1;

Step 6 Run the following command to query the value of the **allow_drop_detached** parameter:

SELECT * FROM system.settings WHERE name = 'allow_drop_detached';



```
server-2110081635-0001 :) SELECT *  FROM system.settings  WHERE name = 'allow_drop_detached';

SELECT *
FROM system.settings
WHERE name = 'allow_drop_detached'

Query id: 8211d1ff-5717-49af-929f-8e4170c6e1d1

┌─name────────────────┬─value─┬─changed─┬─description─────────────────────────────────────────────┬─min──┬─max──┬─readonly─┬─type─┐
│ allow_drop_detached │ 1     │       1 │ Allow ALTER TABLE ... DROP DETACHED PART[ITION] ... queries │ NULL │ NULL │        0 │ Bool │
└─────────────────────┴───────┴─────────┴─────────────────────────────────────────────────────────┴──────┴──────┴──────────┴──────┘

1 rows in set. Elapsed: 0.004 sec.
```

Step 7 Run the **q;** command to exit the ClickHouse client.

----End

14.11.35 How Do I Do If an Alarm Indicating Insufficient Memory Is Reported During Spark Task Execution?

Symptom

When a Spark task is executed, an alarm indicating insufficient memory is reported. The alarm ID is 18022. As a result, no available memory can be used.

Procedure

Set the executor parameters in the SQL script to limit the number of cores and memory of an executor.

For example, the configuration is as follows:

```
set hive.execution.engine=spark;
set spark.executor.cores=2;
set spark.executor.memory=4G;
set spark.executor.instances=10;
```

Change the values of the parameters as required.

14.11.36 How Do I Do If ClickHouse Consumes Excessive CPU Resources?

Symptom

A user performs a large number of update operations using ClickHouse. This operation on a ClickHouse consumes a large number of resources. In addition, the operation will be executed again if it fails. As a result, retries of those failed operations occupy too many CPU resources.

Procedure

Delete existing data from ZooKeeper and release delete the update statement.

14.11.37 How Do I Enable the Map Type on ClickHouse?

Step 1 Log in to the active Master node as user **root**.

Step 2 Run the following command to modify the **/opt/Bigdata/components/current/ClickHouse/configurations.xml** configuration file to enable user parameter customization:

vim /opt/Bigdata/components/current/ClickHouse/configurations.xml

Change **hidden** to **advanced**, as shown in the following information in bold. Then save the configuration and exit.

```
<property type="hidden" scope="all" classification="Customization"
classdesc="RESID_CLICKHOUSE_CONF_0056">
        <name>_clickhouse.custom_content.key</name>
        <value>_user-xml-content</value>
</property>
<property type="advanced" scope="all" classification="Customization"
classdesc="RESID_CLICKHOUSE_CONF_0056">
       <name>_user-xml-content</name>
       <value vType="text" checker="clickhouse.xmlformat">&lt;yandex&gt;&lt;/yandex&gt;</value>
       <description>RESID_CLICKHOUSE_CONF_0025</description>
</property>
```

Step 3 Run the following commands to switch to user **omm** and restart the controller service:

su - omm

sh /opt/Bigdata/om-server/om/sbin/restart-controller.sh

Step 4 Log in to FusionInsight Manager, choose **Cluster** > **Services** > **ClickHouse**. On the page that is displayed, click the **Configurations** tab then the **All Configurations** sub-tab. Click **ClickHouseServer(Role)** > **Customization**, and add the following content to the **_user-xml-content** configuration item in the right pane:

```
<yandex>
  <profiles>
    <default>
   <allow_experimental_map_type>1</allow_experimental_map_type>
    </default>
  </profiles>
</yandex>
```

Step 5 Click **Save**.

Step 6 Choose **Cluster** > **Services** > **ClickHouse**. In the upper right corner, choose **More** > **Restart Service** to restart the ClickHouse service.

----End

14.11.38 A Large Number of OBS APIs Are Called When Spark SQL Accesses Hive Partitioned Tables

Symptom

When Spark SQL is used to access Hive partitioned tables stored in OBS, the acces speed is slow and a large number of OBS query APIs are called.

Example SQL:

```
select a,b,c from test where b=xxx
```

Fault Locating

According to the configuration, the task should scan only the partition whose b is *xxx*. However, the task logs show that the task scans all partitions and then calculates the data whose b is *xxx*. As a result, the task calculation is slow. In addition, a large number of OBS requests are sent because all files need to be scanned.

By default, the execution plan optimization based on partition statistics is enabled on MRS, which is equivalent to automatic execution of Analyze Table. (The default configuration method is to set **spark.sql.statistics.fallBackToHdfs** to **true**. You can set this parameter to **false**.) After this function is enabled, table partition statistics are scanned during SQL execution and used as cost estimation in the execution plan. For example, small tables identified during cost evaluation are broadcast to each node in the memory for join operations, significantly reducing shuffle time. This function greatly optimizes performance in join scenarios, but increases the number of OBS calls.

Procedure

Set the following parameter in Spark SQL and then run the SQL statement:

```
set spark.sql.statistics.fallBackToHdfs=false;
```

Alternatively, run the **--conf** command to set this parameter to **false** before startup.

```
--conf spark.sql.statistics.fallBackToHdfs=false
```

14.12 API

14.12.1 How Do I Configure the node_id Parameter When Using the API for Adjusting Cluster Nodes?

When you use the API for adjusting cluster nodes, the value of **node_id** is fixed to **node_orderadd**.

14.13 Cluster Management

14.13.1 How Do I View All Clusters?

You can view all MRS clusters on the **Clusters** page. You can view clusters in different status.

- **Active Clusters**: all clusters except clusters in **Failed** and **Terminated** states.
- **Cluster History**: clusters in the **Terminated** state. Only the clusters terminated within the last six months are displayed. If you want to view clusters terminated more than six months ago, contact technical support engineers.
- **Failed Tasks**: tasks in **Failed** state. The failed tasks include the following:
  - Tasks failed to create clusters
  - Tasks failed to terminate clusters
  - Tasks failed to scale out clusters
  - Tasks failed to scale in clusters

14.13.2 How Do I View Log Information?

You can view operation logs of clusters and jobs on the **Operation Logs** page. The MRS operation logs record the following operations:

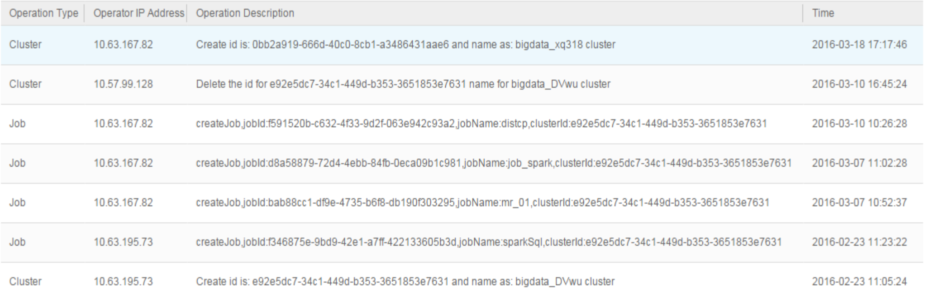
- Cluster operations

- Create, terminate, and scale out or in clusters
- Create directories and delete directories or files

- Job operations: Create, stop, and delete jobs
- Data operations: IAM user tasks, add users, and add user groups

Figure 14-7 shows the operation logs.

Figure 14-7 Log information



Operation Type	Operator IP Address	Operation Description	Time
Cluster	10.63.167.82	Create id is: 0bb2a919-666d-40c0-8cb1-a3486431aae6 and name as: bigdata_xq318 cluster	2016-03-18 17:17:46
Cluster	10.57.99.128	Delete the id for e92e5dc7-34c1-449d-b353-3651853e7631 name for bigdata_DVwu cluster	2016-03-10 16:45:24
Job	10.63.167.82	createJob,jobId:f591520b-c632-4f33-9d2f-063e942c93a2,jobName:distcp,clusterId:e92e5dc7-34c1-449d-b353-3651853e7631	2016-03-10 10:26:28
Job	10.63.167.82	createJob,jobId:d8a58879-72d4-4ebb-84fb-0eca09b1c981,jobName:job_spark,clusterId:e92e5dc7-34c1-449d-b353-3651853e7631	2016-03-07 11:02:28
Job	10.63.167.82	createJob,jobId:bab88cc1-df9e-4735-b6f8-db190f303295,jobName:mr_01,clusterId:e92e5dc7-34c1-449d-b353-3651853e7631	2016-03-07 10:52:37
Job	10.63.195.73	createJob,jobId:f346875e-9bd9-42e1-a7ff-422133605b3d,jobName:sparkSql,clusterId:e92e5dc7-34c1-449d-b353-3651853e7631	2016-02-23 11:23:22
Cluster	10.63.195.73	Create id is: e92e5dc7-34c1-449d-b353-3651853e7631 and name as: bigdata_DVwu cluster	2016-02-23 11:05:24

14.13.3 How Do I View Cluster Configuration Information?

- After a cluster is created, click the cluster name on the MRS console. On the page displayed, you can view basic configuration information about the cluster. The instance specifications and node capacity determine the data analysis and processing capability. Higher instance specifications and larger capacity enable faster data processing at a higher cost.
- On the basic information page, click **Access Manager** to access the MRS cluster management page. On MRS Manager, you can view and handle alarms, and modify cluster configuration.

14.13.4 How Do I Install Kafka and Flume in an MRS Cluster?

You cannot install the Kafka and Flume components for a created cluster of MRS 3.1.0 or earlier. Kafka and Flume are components for a streaming cluster. To install Kafka and Flume, create a streaming or hybrid cluster, and install Kafka and Flume.

14.13.5 How Do I Stop an MRS Cluster?

To stop an MRS cluster, stop each node in the cluster on the ECS. Click the name of each node on the **Nodes** tab page to go to the **Elastic Cloud Server** page and click **Stop**.

14.13.6 Can I Expand Data Disk Capacity for MRS?

You can expand data disk capacity for MRS during off-peak hours.

Expand the EVS disk capacity, and then log in to the ECS and expand the partitions and file system. MRS nodes are installed using public images and support the capacity expansion of in-use EVS disks.

14.13.7 Can I Add Components to an Existing Cluster?

You cannot add or remove any component to and from a created cluster of MRS 3.1.0. However, you can create an MRS cluster that contains the required components.

14.13.8 Can I Delete Components Installed in an MRS Cluster?

You cannot delete any component from a created MRS cluster of MRS 3.1.0. If a component is not required, log in to MRS Manager and stop the component on the **Services** page.

14.13.9 Can I Change MRS Cluster Nodes on the MRS Console?

You cannot change MRS cluster nodes on the MRS console. You are also advised not to change MRS cluster nodes on the ECS console. Manually stopping or deleting an ECS, modifying or reinstalling the ECS OS, or modifying ECS specifications for a cluster node on the ECS console will affect the cluster stability.

If an ECS is deleted, the ECS OS is modified or reinstalled, or the ECS specifications are modified on the ECS console, MRS will automatically identify and delete the node. You can log in to the MRS console and restore the deleted node through scale-out. Do not perform operations on the nodes that are being scaled out.

14.13.10 How Do I Shield Cluster Alarm/Event Notifications?

1. Log in to the MRS console.
2. Click the name of the cluster.
3. On the page displayed, choose **Alarms** > **Notification Rules**.
4. Locate the row that contains the rule you want to modify, click **Edit** in the **Operation** column, and deselect the alarm or event severity levels.
5. Click **OK**.

14.13.11 Why Is the Resource Pool Memory Displayed in the MRS Cluster Smaller Than the Actual Cluster Memory?

In an MRS cluster, MRS allocates 50% of the cluster memory to Yarn by default. You manage Yarn nodes logically by resource pool. Therefore, the total memory of the resource pool displayed in the cluster is only 50% of the total memory of the cluster.

14.13.12 How Do I Configure the knox Memory?

Step 1 Log in to a Master node of the cluster as user **root**.

Step 2 Run the following command on the Master node to open the **gateway.sh** file:

```
su omm
```

```
vim /opt/knox/bin/gateway.sh
```

Step 3 Change **APP_MEM_OPTS=""** to **APP_MEM_OPTS="-Xms256m -Xmx768m"**, save the file, and exit.

Step 4 Run the following command on the Master node to restart the knox process:

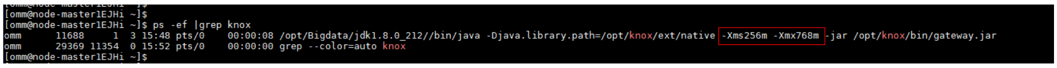
sh /opt/knox/bin/gateway.sh stop

sh /opt/knox/bin/gateway.sh start

Step 5 Repeat the preceding steps on each Master node.

Step 6 Run the **ps -ef |grep knox** command to check the configured memory.

Figure 14-8 knox memory



```
[omm@node-master1EJHi ~]$
[omm@node-master1EJHi ~]$ ps -ef |grep knox
omm      11688     1  3 15:48 pts/0    00:00:08 /opt/Bigdata/jdk1.8.0_212//bin/java -Djava.library.path=/opt/knox/ext/native -Xms256m -Xmx768m -jar /opt/knox/bin/gateway.jar
omm      29369 11354  0 15:52 pts/0    00:00:00 grep --color=auto knox
[omm@node-master1EJHi ~]$
```

----End

14.13.13 What Is the Python Version Installed for an MRS Cluster?

Log in to a Master node as user **root** and run the **Python3** command to query the Python version.

14.13.14 How Do I View the Configuration File Directory of Each Component?

The configuration file paths of commonly used components are as follows:

Component	Configuration File Directory
ClickHouse	*Client installation directory*/**ClickHouse/clickhouse/config**
Flink	*Client installation directory*/**Flink/flink/conf**
Flume	*Client installation directory*/**fusioninsight-flume-***xxx***/conf**
HBase	*Client installation directory*/**HBase/hbase/conf**
HDFS	*Client installation directory*/**HDFS/hadoop/logs/hadoop.log**
Hive	*Client installation directory*/**Hive/config**
Hudi	*Client installation directory*/**Hudi/hudi/conf**
Kafka	*Client installation directory*/**Kafka/kafka/config**

Component	Configuration File Directory
Loader	- *Client installation directory*/**Loader/loader-tools-xxx/loader-tool/conf** - *Client installation directory*/**Loader/loader-tools-xxx/schedule-tool/conf** - *Client installation directory*/**Loader/loader-tools-xxx/shell-client/conf** - *Client installation directory*/**Loader/loader-tools-xxx/sqoop-shell/conf**
Oozie	*Client installation directory*/**Oozie/oozie-client-***xxx***/conf**
Spark2x	*Client installation directory/Spark2x/spark/conf*
Yarn	*Client installation directory/Yarn/config*
ZooKeeper	*Client installation directory/Zookeeper/zookeeper/conf*

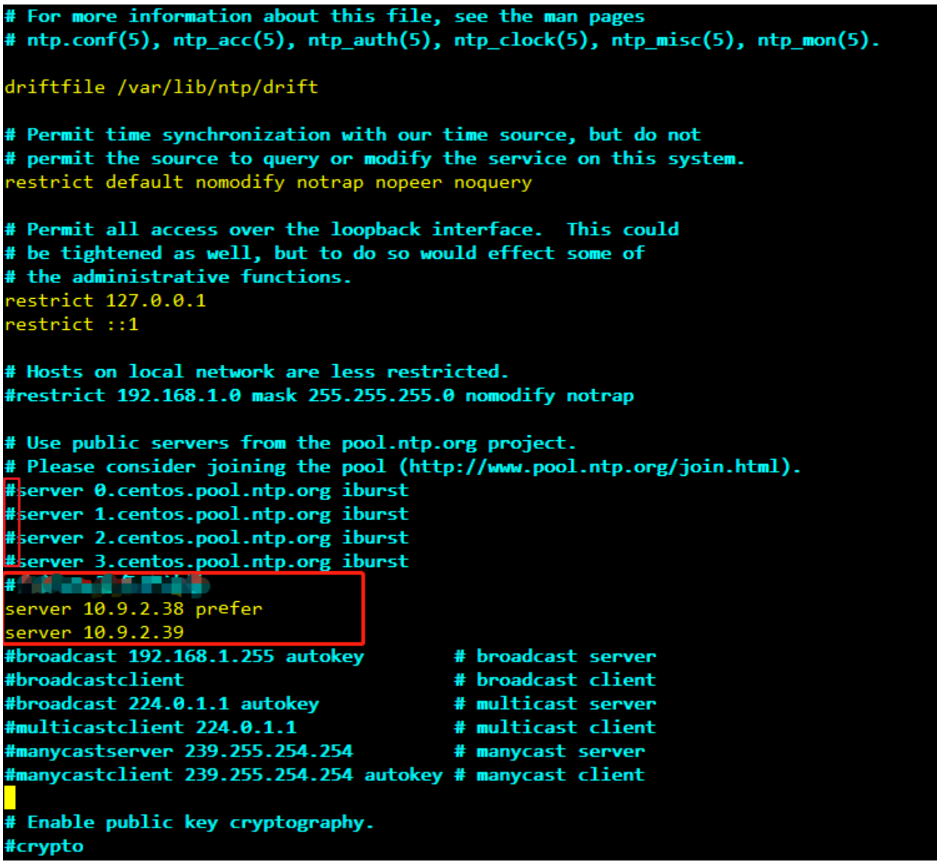
14.13.15 How Do I Do If the Time on MRS Nodes Is Incorrect?

- If the time on a node inside the cluster is incorrect, log in to the node and rectify the fault from 2.
- If the time on a node inside the cluster is different from that on a node outside the cluster, log in to the node and rectify the fault from 1.

1. Run the **vi /etc/ntp.conf** command to edit the NTP client configuration file, add the IP addresses of the master node in the MRS cluster, and comment out the IP address of other servers.

```
server master1_ip prefer
server master2_ip
```

Figure 14-9 Adding the master node IP addresses



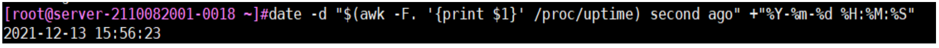
```
# For more information about this file, see the man pages
# ntp.conf(5), ntp_acc(5), ntp_auth(5), ntp_clock(5), ntp_misc(5), ntp_mon(5).

driftfile /var/lib/ntp/drift

# Permit time synchronization with our time source, but do not
# permit the source to query or modify the service on this system.
restrict default nomodify notrap nopeer noquery

# Permit all access over the loopback interface.  This could
# be tightened as well, but to do so would effect some of
# the administrative functions.
restrict 127.0.0.1
restrict ::1

# Hosts on local network are less restricted.
#restrict 192.168.1.0 mask 255.255.255.0 nomodify notrap

# Use public servers from the pool.ntp.org project.
# Please consider joining the pool (http://www.pool.ntp.org/join.html).
#server 0.centos.pool.ntp.org iburst
#server 1.centos.pool.ntp.org iburst
#server 2.centos.pool.ntp.org iburst
#server 3.centos.pool.ntp.org iburst
#[redacted]
server 10.9.2.38 prefer
server 10.9.2.39
#broadcast 192.168.1.255 autokey        # broadcast server
#broadcastclient                        # broadcast client
#broadcast 224.0.1.1 autokey            # multicast server
#multicastclient 224.0.1.1              # multicast client
#manycastserver 239.255.254.254         # manycast server
#manycastclient 239.255.254.254 autokey # manycast client

# Enable public key cryptography.
#crypto
```

2. Run the **service ntpd stop** command to stop the NTP service.
3. Run the **/usr/sbin/ntpdate** *IP address of the active master node* command to manually synchronize time.
4. Run the **service ntpd start** or **systemctl restart ntpd** command to start the NTP service.
5. Run the **ntpstat** command to check the time synchronization result:

14.13.16 How Do I Query the Startup Time of an MRS Node?

Log in to the target node and run the following command to query the startup time:

date -d "$(awk -F. '{print $1}' /proc/uptime) second ago" +"%Y-%m-%d %H:%M:%S"

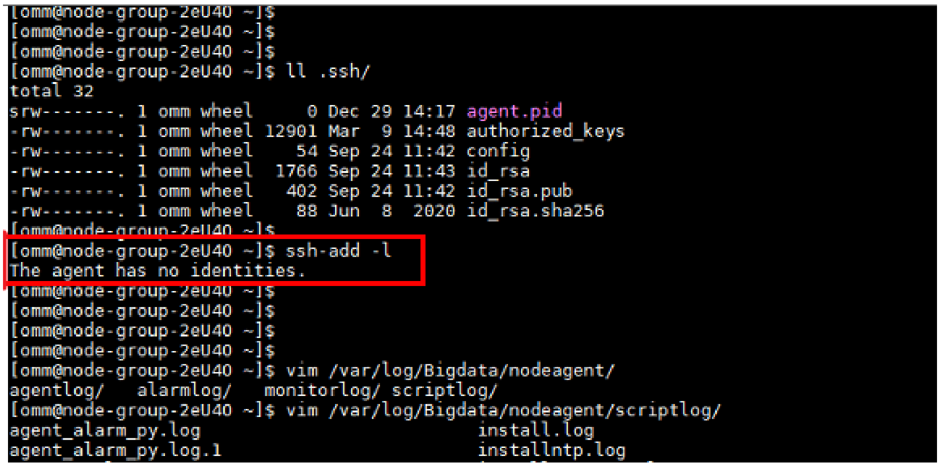


```
[root@server-2110082001-0018 ~]#date -d "$(awk -F. '{print $1}' /proc/uptime) second ago" +"%Y-%m-%d %H:%M:%S"
2021-12-13 15:56:23
```

14.13.17 How Do I Do If Trust Relationships Between Nodes Are Abnormal?

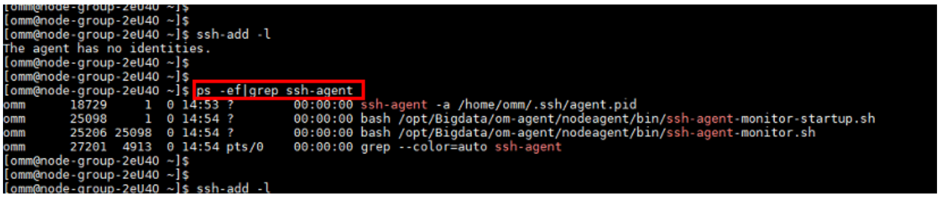
If "ALM-12066 Inter-Node Mutual Trust Fails" is reported on Manager or there is no SSH trust relationship between nodes, rectify the fault by performing the following operations:

1. Run the **ssh-add -l** command on both nodes of the trusted cluster to check whether there are identities.



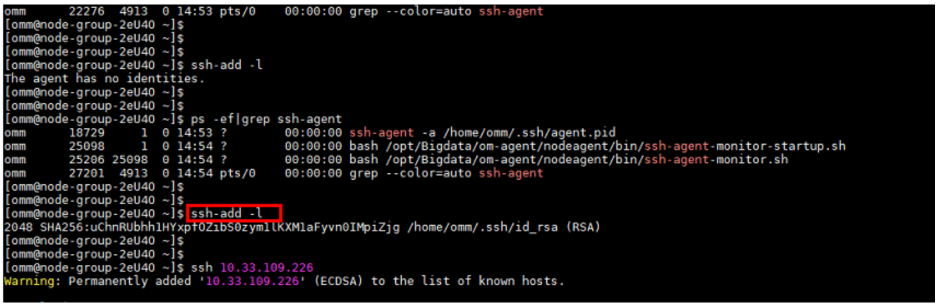
```
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ll .ssh/
total 32
srw-------. 1 omm wheel     0 Dec 29 14:17 agent.pid
-rw-------. 1 omm wheel 12901 Mar  9 14:48 authorized_keys
-rw-------. 1 omm wheel    54 Sep 24 11:42 config
-rw-------. 1 omm wheel  1766 Sep 24 11:43 id_rsa
-rw-------. 1 omm wheel   402 Sep 24 11:42 id_rsa.pub
-rw-------. 1 omm wheel    88 Jun  8  2020 id_rsa.sha256
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh-add -l
The agent has no identities.
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ vim /var/log/Bigdata/nodeagent/
agentlog/   alarmlog/   monitorlog/ scriptlog/
[omm@node-group-2eU40 ~]$ vim /var/log/Bigdata/nodeagent/scriptlog/
agent_alarm_py.log                           install.log
agent_alarm_py.log.1                         installntp.log
```

2. If no identities are displayed, run the **ps -ef|grep ssh-agent** command to find the ssh-agent process, kill the process, and wait for the process to automatically restart.



```
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh-add -l
The agent has no identities.
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ps -ef|grep ssh-agent
omm      18729     1  0 14:53 ?        00:00:00 ssh-agent -a /home/omm/.ssh/agent.pid
omm      25098     1  0 14:54 ?        00:00:00 bash /opt/Bigdata/om-agent/nodeagent/bin/ssh-agent-monitor-startup.sh
omm      25206 25098  0 14:54 ?        00:00:00 bash /opt/Bigdata/om-agent/nodeagent/bin/ssh-agent-monitor.sh
omm      27201  4913  0 14:54 pts/0    00:00:00 grep --color=auto ssh-agent
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh-add -l
```

3. Run the **ssh-add -l** command to check whether the identities have been added. If yes, manually run the **ssh** command to check whether the trust relationship is normal.



```
omm      22276  4913  0 14:53 pts/0    00:00:00 grep --color=auto ssh-agent
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh-add -l
The agent has no identities.
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ps -ef|grep ssh-agent
omm      18729     1  0 14:53 ?        00:00:00 ssh-agent -a /home/omm/.ssh/agent.pid
omm      25098     1  0 14:54 ?        00:00:00 bash /opt/Bigdata/om-agent/nodeagent/bin/ssh-agent-monitor-startup.sh
omm      25206 25098  0 14:54 ?        00:00:00 bash /opt/Bigdata/om-agent/nodeagent/bin/ssh-agent-monitor.sh
omm      27201  4913  0 14:54 pts/0    00:00:00 grep --color=auto ssh-agent
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh-add -l
2048 SHA256:uChnRUbhh1HYxpf0ZibS0zym1lKXMlaFyvn0IMpiZjg /home/omm/.ssh/id_rsa (RSA)
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh 10.33.109.226
Warning: Permanently added '10.33.109.226' (ECDSA) to the list of known hosts.
```

4. If identities exist, check whether the **authorized_keys** file in the **/home/omm/.ssh** directory contains the information in the **id_rsa.pub** file in the **/home/omm/.ssh** of the peer node. If no, manually add the information about the peer node.
5. Check whether the permissions on the files in **/home/omm/.ssh** directory are correct.
6. Check the **/var/log/Bigdata/nodeagent/scriptlog/ssh-agent-monitor.log** file.
7. If the **home** directory of user **omm** is deleted, contact MRS support personnel.

14.13.18 How Do I Adjust the Memory Size of the manager-executor Process?

Symptom

The **manager-executor** process runs either on the Master1 or Master2 node in the MRS cluster in active/standby mode. This process is used to encapsulate the MRS management and control plane's operations on the MRS cluster, such as job submission, heartbeat reporting, certain alarm reporting, as well as cluster creation, scale-out, and scale-in. When you submit jobs on the MRS management and control plane, the Executor memory may become insufficient as the tasks increase or the number of concurrent tasks increases. As a result, the CPU usage is high and the Executor process experiences out-of-memory (OOM) errors.

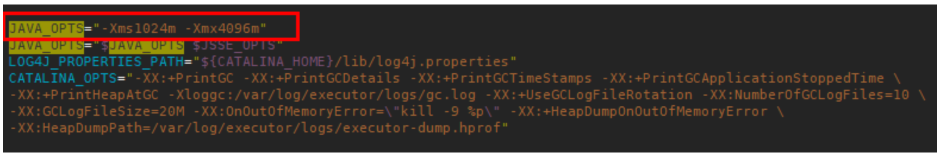
Procedure

1. Log in to either the Master1 or Master2 node as user **root** and run the following command to switch to user **omm**:

   su - omm

2. Run the following command to modify the **catalina.sh** script. Specifically, search for **JAVA_OPTS** in the script, find the configuration items similar to **JAVA_OPTS="-Xms1024m -Xmx4096m**, and change the values of the items to desired ones, and save the modification.

   vim /opt/executor/bin/catalina.sh

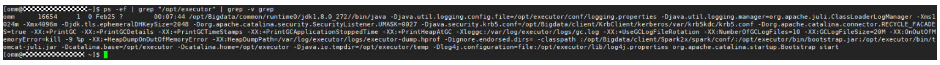
   

   ```
   JAVA_OPTS="-Xms1024m -Xmx4096m"
   JAVA_OPTS="$JAVA_OPTS $JSSE_OPTS"
   LOG4J_PROPERTIES_PATH="${CATALINA_HOME}/lib/log4j.properties"
   CATALINA_OPTS="-XX:+PrintGC -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -XX:+PrintGCApplicationStoppedTime \
   -XX:+PrintHeapAtGC -Xloggc:/var/log/executor/logs/gc.log -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 \
   -XX:GCLogFileSize=20M -XX:OnOutOfMemoryError=\"kill -9 %p\" -XX:+HeapDumpOnOutOfMemoryError \
   -XX:HeapDumpPath=/var/log/executor/logs/executor-dump.hprof"
   ```

3. The **manager-executor** process only runs on either the Master1 or Master2 node in active/standby mode. Check whether it exists on the node before restarting it.

   a. Log in to the Master1 and Master2 nodes and run the following command to check whether the process exists. If any command output is displayed, the process exists.

      ps -ef | grep "/opt/executor" | grep -v grep

      ```
      [omm@[redacted]~]$ ps -ef | grep "/opt/executor" | grep -v grep
      omm      16654     1  0 Feb25 ?        00:07:44 /opt/Bigdata/common/runtime0/jdk1.8.0_272//bin/java -Djava.util.logging.config.file=/opt/executor/conf/logging.properties -Djava.util.logging.manager=org.apache.juli.ClassLoaderLogManager -Xms1
      024m -Xmx4096m -Djdk.tls.ephemeralDHKeySize=2048 -Dorg.apache.catalina.security.SecurityListener.UMASK=0027 -Djava.security.krb5.conf=/opt/Bigdata/client/KrbClient/kerberos/var/krb5kdc/krb5.conf -Dorg.apache.catalina.connector.RECYCLE_FACADE
      S=true -XX:+PrintGC -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -XX:+PrintGCApplicationStoppedTime -XX:+PrintHeapAtGC -Xloggc:/var/log/executor/logs/gc.log -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=20M -XX:OnOutOfM
      emoryError=kill -9 %p -XX:+HeapDumpOnOutOfMemoryError -XX:HeapDumpPath=/var/log/executor/logs/executor-dump.hprof -Dignore.endorsed.dirs= -classpath :/opt/Bigdata/client/Spark2x/spark/conf/:/opt/executor/bin/bootstrap.jar:/opt/executor/bin/t
      omcat-juli.jar -Dcatalina.base=/opt/executor -Dcatalina.home=/opt/executor -Djava.io.tmpdir=/opt/executor/temp -Dlog4j.configuration=file:/opt/executor/lib/log4j.properties org.apache.catalina.startup.Bootstrap start
      [omm@[redacted] ~]$
      ```

   b. Run the following command to restart the process:

      sh /opt/executor/bin/shutdown.shsh /opt/executor/bin/startup.sh

14.14 Kerberos Usage

14.14.1 How Do I Change the Kerberos Authentication Status of a Created MRS Cluster?

You cannot change the Kerberos service after an MRS cluster is created.

14.14.2 What Are the Ports of the Kerberos Authentication Service?

The Kerberos authentication service uses ports 21730 (TCP), 21731 (TCP/UDP), and 21732 (TCP/UDP).

14.14.3 How Do I Deploy the Kerberos Service in a Running Cluster?

The MRS cluster does not support customized Kerberos installation and deployment, and the Kerberos authentication cannot be set up between components. To enable Kerberos authentication, you need to create a cluster with Kerberos enabled and migrate data.

14.14.4 How Do I Access Hive in a Cluster with Kerberos Authentication Enabled?

1. Log in to the master node in the cluster as user **root**.
2. Run the following command to configure environment variables:

   source /opt/client/bigdata_env
3. If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user:

   kinit *MRS cluster user*

   Example: **kinit hiveuser**

   The current user must have the permission to create Hive tables..
4. Run the client command of the Hive component.

   beeline
5. Run the Hive command in Beeline, for example:

   create table test_obs(a int, b string) row format delimited fields terminated by "," stored as textfile location "obs://test_obs";
6. Press **Ctrl+C** to exit the Hive Beeline.

14.14.5 How Do I Access Presto in a Cluster with Kerberos Authentication Enabled?

1. Log in to the Master node in the cluster as user **root**.
2. Run the following command to configure environment variables:

   source /opt/client/bigdata_env
3. Access Presto in a cluster with Kerberos authentication enabled.
   a. Log in to MRS Manager and create a role with the **Hive Admin Privilege** permission, for example, **prestorerole**.
   b. Create a user, for example, **presto001**, who belongs to the **Presto** and **Hive** groups, and bind the user to the role created in **3.a**.
   c. Authenticate user **presto001**.

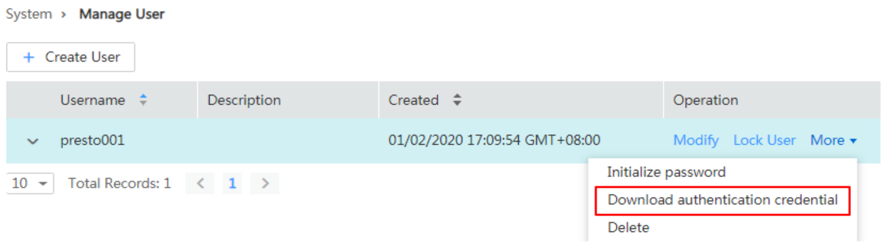
      kinit presto001

d. Download the user authentication credential.

- Operations on MRS Manager:

  Log in to MRS Manager, choose **System** > **Manage User**. Locate the user, and choose **More** > **Download authentication credential**.

  Figure 14-10 Downloading the Presto user authentication credential



Username	Description	Created	Operation
presto001		01/02/2020 17:09:54 GMT+08:00	Modify Lock User More

- Operations on FusionInsight Manager:

  Log in to FusionInsight Manager, choose **System** > **Permission** > **User**. On the displayed page, locate the row that contains the user, choose **More** > **Download Authentication Credential**.

e. Decompress the downloaded user credential file, and save the obtained **krb5.conf** and **user.keytab** files to the client directory, for example, /opt/**client/Presto/**.

f. Run the following command to obtain the user principal:

   klist -kt /opt/client/Presto/user.keytab

g. Run the following command to connect to the Presto Server of the cluster:

   presto_cli.sh --krb5-config-path *{krb5.conf file path}* **--krb5-principal** *{User's principal}* **--krb5-keytab-path** *{user.keytab file path}* **--user** *{presto username}*

- **krb5.conf** *file path*: file path set in 3.e, for example, **/opt/client/Presto/krb5.conf**.
- **user.keytab** *file path*: file path set in 3.e, for example, **/opt/client/Presto/user.keytab**.
- *User's principal*: principal obtained in 3.f.
- *presto username*: user created in 3.b, for example, **presto001**.

   Example: **presto_cli.sh --krb5-config-path /opt/client/Presto/krb5.conf --krb5-principal prest001@xxx_xxx_xxx_xxx.COM --krb5-keytab-path /opt/client/Presto/user.keytab --user presto001**

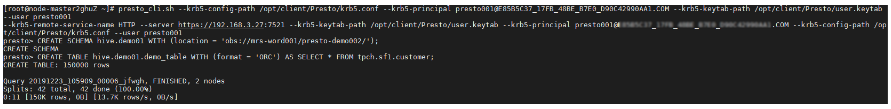
h. On the Presto client, run the following statement to create a schema:

   CREATE SCHEMA hive.demo01 WITH (location = 'obs://presto-demo002/');

i. Create a table in the schema. The table data is stored in the OBS bucket, as shown in the following example:

   CREATE TABLE hive.demo01.demo_table WITH (format = 'ORC') AS SELECT * FROM tpch.sf1.customer;

Figure 14-11 Return result



```
[root@node-master2ghuZ ~]# presto_cli.sh --krb5-config-path /opt/client/Presto/krb5.conf --krb5-principal presto001@E85B5C37_17FB_48BE_B7E0_D90C42990AA1.COM --krb5-keytab-path /opt/client/Presto/user.keytab
--user presto001
--krb5-remote-service-name HTTP --server https://192.168.3.27:7521 --krb5-keytab-path /opt/client/Presto/user.keytab --krb5-principal presto001@[redacted].COM --krb5-config-path /opt/client/Presto/krb5.conf --user presto001
presto> CREATE SCHEMA hive.demo01 WITH (location = 'obs://mrs-word001/presto-demo002/');
CREATE SCHEMA
presto> CREATE TABLE hive.demo01.demo_table WITH (format = 'ORC') AS SELECT * FROM tpch.sf1.customer;
CREATE TABLE: 150000 rows

Query 20191223_105909_00006_jfwgh, FINISHED, 2 nodes
Splits: 42 total, 42 done (100.00%)
0:11 [150K rows, 0B] [13.7K rows/s, 0B/s]
```

j. Run **exit** to exit the client.

14.14.6 How Do I Access Spark in a Cluster with Kerberos Authentication Enabled?

1. Log in to the master node in the cluster as user **root**.
2. Run the following command to configure environment variables:

   source /opt/client/bigdata_env
3. If the Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user.

   kinit *MRS cluster user*

   Example:

   If the development user is a machine-machine user, run **kinit -kt user.keytab sparkuser**.

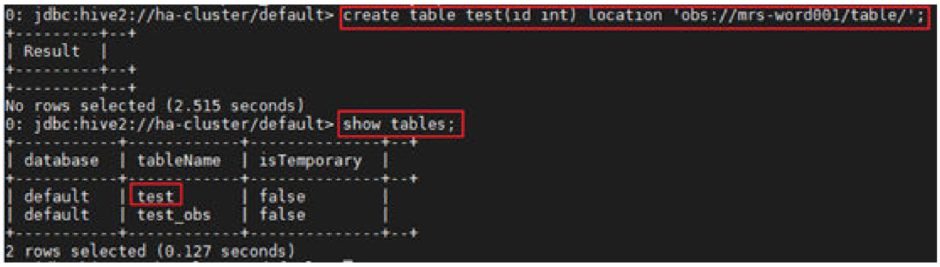
   If the development user is a human-machine user, run **kinit sparkuser**.
4. Run the following command to connect to Spark Beeline:

   spark-beeline
5. Run commands on Spark Beeline. For example, create the table **test** in the **obs://mrs-word001/table/** directory.

   create table test(id int) location 'obs://mrs-word001/table/';
6. Run the following command to query all tables. If table **test** is displayed in the command output, OBS access is successful.

   show tables;

   Figure 14-12 Returned table name

```
0: jdbc:hive2://ha-cluster/default> create table test(id int) location 'obs://mrs-word001/table/';
+---------+--+
| Result  |
+---------+--+
+---------+--+
No rows selected (2.515 seconds)
0: jdbc:hive2://ha-cluster/default> show tables;
+-----------+------------+--------------+--+
| database  | tableName  | isTemporary  |
+-----------+------------+--------------+--+
| default   | test       | false        |
| default   | test_obs   | false        |
+-----------+------------+--------------+--+
2 rows selected (0.127 seconds)
```

7. Press **Ctrl+C** to exit Spark Beeline.

14.14.7 How Do I Prevent Kerberos Authentication Expiration?

- Java applications:

  Before connecting to HBase, HDFS, or other big data components, call loginUserFromKeytab() to create a UGI. Then, start a scheduled thread to periodically check whether the Kerberos Authentication expires. Log in to the system again before the Kerberos Authentication expires.

```
private static void startCheckKeytabTgtAndReloginJob() {
//The credential is checked every 10 minutes, and updated before the expiration time.
        ThreadPool.updateConfigThread.scheduleWithFixedDelay(() -> {
            try {
                UserGroupInformation.getLoginUser().checkTGTAndReloginFromKeytab();
                logger.warn("get tgt:{}", UserGroupInformation.getLoginUser().getTGT());
                logger.warn("Check Kerberos Tgt And Relogin From Keytab Finish.");
            } catch (IOException e) {
                logger.error("Check Kerberos Tgt And Relogin From Keytab Error", e);
            }
        }, 0, 10, TimeUnit.MINUTES);
        logger.warn("Start Check Keytab TGT And Relogin Job Success.");
    }
```

- Tasks executed in shell mode:

  a. Run the **kinit** command to authenticate the user.

  b. Create a scheduled task of the operating system or any other scheduled task to run the **kinit** command to authenticate the user periodically.

  c. Submit jobs to execute big data tasks.

- Spark jobs:

  If you submit jobs using spark-shell, spark-submit, or spark-sql, you can specify **Keytab** and **Principal** in the command to perform authentication and periodically update the login credential and authorization tokens to prevent authentication expiration.

  Example:

  spark-shell --principal spark2x/hadoop.<*System domain name*>@<*System domain name*> **--keytab ${BIGDATA_HOME}/FusionInsight_Spark2x_8.1.0.1/install/FusionInsight-Spark2x-2.4.5/keytab/spark2x/SparkResource/spark2x.keytab --master yarn**

14.15 Metadata Management

14.15.1 Where Can I View Hive Metadata?

- If Hive metadata is stored in GaussDB of an MRS cluster, log in to the master DBServer node of the cluster, switch to user **omm**, and run the **gsql -p 20051 -U {USER} -W {PASSWD} -d hivemeta** command to view the metadata.
- If Hive metadata is stored in an external relational database, perform the following steps:

  a. On the cluster **Dashboard** page, click **Manage** on the right of **Data Connection**.

  b. On the displayed page, obtain the value of **Data Connection ID**.

  c. On the MRS console, click **Data Connections**.

  d. In the data connection list, locate the data connection based on the data connection ID obtained in **b**.

  e. Click **Edit** in the **Operation** column of the data connection.

     The **RDS Instance** and **Database** indicate the relational database in which the Hive metadata is stored.

15 Troubleshooting

Released On	What's New
2021-06-30	This issue is the fifth official release. Released MRS 3.1.0-LTS.1. For details, see **Creating a Custom Cluster**.
2020-10-24	This issue is the fourth official release. Added the following sections: ● **Methods of Creating MRS Clusters** ● **Quick Creation of a Hadoop Analysis Cluster** ● **Quick Creation of an HBase Analysis Cluster** ● **Quick Creation of a Kafka Streaming Cluster** Modified the following sections: ● **Creating a Custom Cluster** ● **Viewing Basic Cluster Information** ● **Viewing Information of a Historical Cluster**
2020-01-17	This issue is the third official release. Modified the following sections: ● **Creating a Custom Cluster** ● **Viewing Basic Cluster Information** ● **Viewing Information of a Historical Cluster**

Released On	What's New
2019-03-20	This issue is the second official issue. Added the following sections: ● **Security Configuration Suggestions for Clusters with Kerberos Authentication Disabled** ● **Authorizing O&M** ● **Authorizing O&M** ● **Adding a Tag to a Cluster** ● **Installing Third-Party Software Using Bootstrap Actions** ● **Restoring Patches for the Isolated Hosts** ● **Rolling Restart** Modified the following sections: ● **Creating a Custom Cluster** ● **Viewing Basic Cluster Information** ● **Configuring an Auto Scaling Rule** ● **Viewing Information of a Historical Cluster**
2018-10-12	This issue is the first official release.

15.1 Accessing the Web Pages

15.1.1 Failed to Access MRS Manager

Symptom

The MRS Manager is not accessible after a cluster is created.

Possible Cause

- MRS can be accessed from an external network only after an EIP is bound to an MRS node.
- Port 9022 is disabled. Add a security group rule to enable the port.

Procedure

Step 1 Log in to the MRS management console, locate the cluster to be accessed in the active cluster list, and click the cluster name.

Step 2 On the node information page, click the name of the node to be accessed, and choose **EIPs** > **Bind EIP**.

Step 3 On the **Bind EIP** page, select a NIC from the **Select NIC** drop-down list, select an EIP from the **Select EIP** list, and click **OK**.

Step 4 After the EIP is bound, enable port 9022 in a security group rule.

Click the **Security Groups** tab. Then, click **Change Security Group**.

You can select an existing security group, or click **Create Security Group** to add a security group rule to enable port 9022 for accessing through the public IP address.

Step 5 After the EIP is added, you can access MRS through **https://***Elastic IP address***: 9022/mrsmanager/**. If the fault still persists, contact technical support for assistance.

----End

15.1.2 Failed to Log In to MRS Manager After the Python Upgrade

Issue

Failed to log in to MRS Manager after Python is upgraded.

Symptom

After Python is upgraded, MRS Manager fails to be accessed using the **admin** account and the correct password.

Possible Cause

When upgrading Python to Python 3.*x*, the user modifies the file directory permission of **openssl**. As a result, the LdapServer service cannot be started, causing a login authentication failure.

Procedure

Step 1 Log in to the Master node in the cluster as user **root**.

Step 2 Run the **chmod 755 /usr/bin/openssl** command to modify the file directory permission of **/usr/bin/openssl** to **755**.

Step 3 Run the **su omm** command to switch to user **omm**.

Step 4 Run the **openssl** command to check whether the **openssl** mode can be entered.

If it can be entered, the permission has been modified successfully. If it cannot be entered, the permission fails to be modified.

If the permission fails to be modified, check whether the command is correct or contact O&M personnel.

Step 5 After the permission is modified, the LdapServer service will be restarted. After the LdapServer service is restarted, log in to MRS Manager again.

----End

Summary and Suggestions

It is recommended that software installed by the user be separated from system software. A system software upgrade may cause compatibility problems.

15.1.3 Failed to Log In to MRS Manager After Changing the Domain Name

Symptom

After changing the domain name, the user cannot log in to MRS Manager through the console, or fails to log in to MRS Manager.

Possible Causes

After the domain name is changed, the **keytab** file of user **executor** is not updated. As a result, the executor process repeatedly performs authentication after the authentication fails, causing memory overflow of the ACS process.

Solution

Step 1 Restart the acs process.

1. Log in to the active management node (master node marked a solid star on the **Nodes** tab of the MRS cluster) as user **root**.
2. Run the following commands to restart the acs process:

   su - omm

   ps -ef|grep =acs (Query the PID of the acs process.)

   kill -9 *PID* (Replace *PID* with the acs process ID to kill the acs process.)
3. Wait for several minutes and run the **ps -ef|grep =acs** command to check whether the acs process is automatically started.

Step 2 Replace the **keytab** file of user **executor**.

1. Log in to MRS Manager and choose **System** > **User**. In the **Operation** column where user **executor** resides, click **Download Authentication Credential**. Decompress the package to obtain the **keytab** file.
2. Log in to the active management node as user **root** and replace the **/opt/executor/webapps/executor/WEB-INF/classes/user.keytab** file with the file obtained in **Step 2.1**.

Step 3 Replace the **keytab** and **conf** files of user **knox**.

1. Log in to MRS Manager and choose **System** > **User**. In the **Operation** column where user **knox** resides, click **Download Authentication Credential**. Decompress the package to obtain the **keytab** and **conf** files.
2. Log in to the active management node as user **root** and replace the **/opt/knox/conf/user.keytab** with the file obtained in **Step 3.1**.
3. Change the **principal** value in the **/opt/knox/conf/krb5JAASLogin.conf** file to the new domain name.
4. Replace the **/opt/knox/conf/krb5.conf** file with the **krb5.conf** file obtained in **Step 3.1**.

Step 4 Back up the original client directory.

mv *{Client directory}* **/opt/client_init**

Step 5 Reinstall the client.

Step 6 Log in to the active and standby management nodes as user **root** and run the following commands to restart the knox process:

su - omm

ps -ef | grep gateway | grep -v grep (Search for the PID of the knox process.)

kill -9 *PID* (Replace *PID* with the ID of the knox process to kill the knox process.)

/opt/knox/bin/restart-knox.sh (Start the knox process.)

Step 7 Log in to the active and standby management nodes as user **root** and run the following commands to restart the executor process:

su - omm

netstat -anp |grep 8181 |grep LISTEN (Search for the PID of the executor process.)

kill -9 *PID* (Replace *PID* with the ID of the executor process to kill the executor process.)

/opt/executor/bin/startup.sh (Start the executor process.)

----End

15.1.4 A Blank Page Is Displayed Upon Login to Manager

Issue

After a user logs in to FusionInsight Manager, the page displayed is blank.

Symptom

After a user logs in to FusionInsight Manager, the page displayed is blank.

Cause Analysis

Login to FusionInsight Manager fails, and the browser cache needs to be cleared.

Procedure

Step 1 Open the browser (using Google Chrome as an example), and press **Ctrl+Shift +Delete**. The dialog box for clearing browsing data is displayed.

Step 2 Select the browsing records to be cleared and click **Clear Data**.

----End

15.1.5 Failed to Download Authentication Credentials When the Username Is Too Long

Issue

In MRS clusters 3.0.2 to 3.1.0, a maximum of 32 characters are allowed in the username when a user is added. However, if the username contains more than 20 characters, the user fails to download the Keytab file, and status code "400 Bad Request" is displayed.

Symptom

In MRS clusters 3.0.2 to 3.1.0, a maximum of 32 characters are allowed in the username when a user is added. However, if the username contains more than 20 characters, the user fails to download the Keytab file, and status code "400 Bad Request" is displayed.

Cause Analysis

The **validate-common-config.xml**, **validate-rule-session.xml**, and **validate-rule-user.xml** configuration files in the **/opt/Bigdata/om-server_*/apache-tomcat-*/webapps/web/WEB-INF/validate** directory of the master node are incorrect and need to be modified.

Procedure

Step 1 Log in to the master node as user **omm** and switch to the **/opt/Bigdata/om-server_*/apache-tomcat-*/webapps/web/WEB-INF/validate** directory.

cd /opt/Bigdata/om-server_*/apache-tomcat-*/webapps/web/WEB-INF/validate

Step 2 Modify the **validate-common-config.xml** file.

vi validate-common-config.xml

Change the **maxLength** value of the username from **32** to **64**.

```
<!-- Username -->
<validators alias="USER_NAME">
      <validator name="RANGE_LENGTH_VALIDATOR" minLength="3"
         maxLength="64" />
      <validator name="REGEXP_VALIDATOR" rule="^[_a-zA-Z0-9\- ]+$"
</validators>
```

Step 3 Modify the **validate-rule-session.xml** file.

vi validate-rule-session.xml

Change the **rule** value from **20** to **64**.

```
 <!-- Download the credentials of the current user -->
 <param_validator url="/api/v2/session/user/keytab/download" method="get"
errorHandler="com.xxx.bigdata.om.web.api.validate.SpecialValidatorErrorHandler" dataPattern="form">
            <!-- Parameter name: File name -->
            <!--Validation rule: userName_13-digit number_keytab.tar; case sensitive-->
            <parameter name="file_name" required="true" errorKey="13-4000005"
errorMessage="RESID_OM_API_SESSION_0013">
```

```
                    <validator name="REGEXP_VALIDATOR" rule="[\-\w ]{3,64}_\d{13}_keytab\.tar" 
caseSensitive="true" />
 </parameter>
```

Step 4 Modify the **validate-rule-user.xml** file.

vi validate-rule-user.xml

Change the **rule** value from **20** to **64**.

```
<!--Download the user credentials  -->
   <param_validator url="/api/v2/permission/users/keytab/download" method="get" 
errorHandler="com.xxx.bigdata.om.web.api.validate.SpecialValidatorErrorHandler" dataPattern="form">
           <!--Mandatory; userName_13-digit number_keytab.tar; case sensitive-->
           <parameter name="file_name" required="true" errorKey="12-4000005" 
errorMessage="RESID_OM_API_AUTHORITY_0005">
                  <validator name="REGEXP_VALIDATOR" rule="[\-\w ]{3,64}_\d{13}_keytab\.tar" 
caseSensitive="true" />
           </parameter>
    </param_validator>
```

Step 5 Restart Tomcat and wait until the startup is successful.

1. Run the following command as user **omm** to query the PID of the Tomcat process:

   ps -ef|grep apache-tomcat

2. Run the **kill -9** *PID* command to forcibly stop the specified Tomcat process. For example:

   kill -9 1203

3. Run the following command to restart Tomcat:

   sh ${BIGDATA_HOME}/om-server/tomcat/bin/startup.sh

Step 6 Download the authentication credentials again.

----End

15.2 Cluster Management

15.2.1 Failed to Reduce Task Nodes

Issue

A user fails to scale in an MRS 2.*x* cluster by reducing the number of task nodes to **0** on the MRS console.

Symptom

When the number of task nodes in an MRS cluster is reduced on the MRS console, the following information is displayed:

This operation is not allowed because the number of instances of NodeManager will be less than the minimum configuration after scale-in, which may cause data loss.

Cause Analysis

The NodeManager service of the core node is stopped. If the number of task nodes is changed to **0**, there will be no NodeManager in the cluster and the Yarn service will be unavailable. Therefore, MRS allows the reduction of task nodes only when the number of NodeManagers is greater than or equal to **1**.

Procedure

Step 1 Choose **Services** > **Yarn** > **Instances**.

- For versions earlier than MRS 2.0.1: Log in to MRS Manager, and choose **Services** > **Yarn** > **Instance**.
- For MRS 2.0.1 or later: Click the cluster name on the MRS console and choose **Components** > **Yarn** > **Instances**.
- For MRS 3.*x* or later: Log in to FusionInsight Manager, choose **Cluster**, click the name of the target cluster, and choose **Services** > **Yarn** > **Instance**.

Step 2 Select the NodeManager instance of the core node, click **More**, and select **Start Instance**.

Step 3 Reduce the number of task nodes on the cluster details page.

1. Click the cluster name, and select the **Nodes** tab.
2. Locate the row that contains the task node group and click **Scale In** in the **Operation** column.
3. Click **OK**. In the displayed dialog box, click **Yes**.

Step 4 After the scale-in is successful, stop NodeManager of the core node if you do not need it.

----End

Summary and Suggestions

You are advised not to stop NodeManager of the core node.

15.2.2 Adding a New Disk to an MRS Cluster

Issue

MRS HBase is unavailable.

Symptom

A high disk usage of the user's host causes service faults.

Cause Analysis

The service becomes unavailable due to insufficient disk capacity of the core node.

Procedure

Step 1 Purchase an EVS disk.

Step 2 Attach the EVS disk.

- If the EVS disk has been attached, go to **Step 6**.
- If an ECS cannot be selected when you attach the EVS disk on the EVS console, go to **Step 3**.

Step 3 Log in to the ECS console and click the name of the ECS to which the new disk is to be attached.

Step 4 On the **Disks** tab, click **Attach Disk**.

Step 5 Select the new disk to be attached and click **OK**.

Step 6 Initialize a Linux data disk.

NOTE

- The mount point directory is the existing DataNode instance ID plus one. For example, if you run the **df -h** command and find that the existing ID is **/srv/BigData/hadoop/data1**, the new mount point is then **/srv/BigData/hadoop/data2**. When initializing a Linux data disk to create a mount point, name the mount point **/srv/BigData/hadoop/data2** and mount a new partition to the mount point. For example:

```
mkdir /srv/BigData/hadoop/data2
mount /dev/xvdb1 /srv/BigData/hadoop/data2
```

About the **/srv/BigData/hadoop/data2** path: Change **/srv/BigData/hadoop/data2** mentioned below according to the following scenarios:

- In 3.*x*: Change it to **/srv/BigData/data2**.
- In versions earlier than 3.*x* : Change it to **/srv/BigData/hadoop/data2**.

Step 7 Run the following command to grant user **omm** the permissions to access the new disk:

chown omm:wheel *New mount point*

Example: **chown omm:wheel /srv/BigData/hadoop/data2**

Step 8 Run the following command to grant the execution permission on the new mount point directory:

chmod 701 *New mount point*

Example: **chmod 701 /srv/BigData/hadoop/data2**

NOTE

In this command, **701** is only an example. Replace it with the value of the existing data disk **data1**.

Step 9 Log in to Manager and add data disks to DataNode and NodeManager instances.

Step 10 Modify the DataNode instance configuration.

MRS Manager: Log in to MRS Manager, choose **Services** > **HDFS** > **Instance**, click the target DataNode instance, and click **Instance Configuration**. On the displayed page, set **Type** to **All**.

FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the desired cluster and choose **Service** > **HDFS** > **Instance**. Click the target DataNode instance, click **Instance Configuration**, and select **All Configurations**.

- Method 1: Manually modify the DataNode instance configuration on the current node.
  - Enter **dfs.datanode.fsdataset.volume.choosing.policy** in the search box and change the parameter value to **org.apache.hadoop.hdfs.server.datanode.fsdataset.AvailableSpaceVolumeChoosingPolicy**.
  - Enter **dfs.datanode.data.dir** in the search box and change the parameter value to **/srv/BigData/hadoop/data1/dn,/srv/BigData/hadoop/data2/dn**.

  If the values of the two parameters have been changed, click **Save Configuration** and select **Restart role instance** to restart the DataNode instance.
- Method 2: Automatically synchronize the DataNode instance configuration on the current node.
  1. Click **Synchronize Configuration** to enable the new configuration for the HDFS service.
  2. After the synchronization is complete, restart the instance for the configuration to take effect.

NOTE

- If HDFS is not used and you want to quickly restart the instance, select **Restart role instance**.
- If a task is using HDFS, you must select rolling restart to prevent data exceptions or task failures.

Step 11 Modify the Yarn NodeManager instance configuration.

MRS Manager: Log in to MRS Manager, choose **Services** > **Yarn** > **Instance**, click the target NodeManager instance, and click **Instance Configuration**. On the displayed page, set **Type** to **All**.

FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the desired cluster and choose **Service** > **Yarn** > **Instance**. Click the target NodeManager instance, click **Instance Configuration**, and select **All Configurations**.

- Method 1: Manually modify the Yarn NodeManager instance configuration on the current node.
  - Enter **yarn.nodemanager.local-dirs** in the search box and change the parameter value to **/srv/BigData/hadoop/data1/nm/localdir,/srv/BigData/hadoop/data2/nm/localdir**.
  - Enter **yarn.nodemanager.log-dirs** in the search box and change the parameter value to **/srv/BigData/hadoop/data1/nm/containerlogs,/srv/BigData/hadoop/data2/nm/containerlogs**.

    If the values of the two parameters have been changed, click **Save Configuration** and select **Restart role instance** to restart the NodeManager instance.
- Method 2: Automatically synchronize the Yarn NodeManager instance configuration on the current node.

a. Click **Synchronize Configuration** to enable the new configuration for the Yarn service.

b. After the synchronization is complete, restart the instance for the configuration to take effect.

NOTE

- If Yarn is not used and you want to quickly restart the instance, select **Restart role instance**.
- If a task is using Yarn, you must select rolling restart to prevent data exceptions or task failures.

Step 12 Check whether the capacity expansion is successful.

MRS Manager: Log in to MRS Manager, choose **Services** > **HDFS** > **Instance**, and click the target DataNode instance.

FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the desired cluster, choose **Service** > **HDFS** > **Instance**, and click the target DataNode instance.

In the **Chart** area, check whether the total disk capacity in real-time monitoring item **DataNode Storage** is increased. If **DataNode Storage** does not exist in the **Chart** area, click **Customize** to add it.

- If the total disk capacity has been increased, the capacity expansion is complete.
- If the total disk capacity does not increase, contact technical support.

Step 13 (Optional) Add data disks to a Kafka instance.

Modify the Kafka instance configuration.

1. Navigate to the parameter settings of the target Kafka Broker node.

   MRS Manager: Log in to MRS Manager, choose **Services** > **Kafka** > **Instance**, click the target Broker instance, and click **Instance Configuration**. On the displayed page, set **Type** to **All**.

   FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the desired cluster and choose **Service** > **Kafka** > **Instance**. Click the target Broker instance, click **Instance Configuration**, and select **All Configurations**.

2. Enter **log.dirs** in the search box, add information about the disks to be added, and use commas (,) to separate them.

   For example, if there is only one existing Kafka data disk and a new one is added, change **/srv/BigData/kafka/data1/kafka-logs** to **/srv/BigData/kafka/data1/kafka-logs,/srv/BigData/kafka/data2/kafka-logs**.

3. Save the configuration and select **Restart role instance** to restart the instance as prompted.

4. Check whether the capacity expansion is successful.

   MRS Manager: Log in to MRS Manager, choose **Services** > **Kafka** > **Instance**, and click the target Broker instance.

   FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the desired cluster, choose **Service** > **Kafka** > **Instance**, and click the target Broker instance.

Check whether the total disk capacity in real-time monitoring item **Capacity of Broker Disks** is increased.

----End

NOTICE

After the disk capacity of a cluster node is expanded, if a new node is added to the cluster, you need to add disks to the new node by referring to the preceding procedure. Otherwise, data may be lost.

Summary and Suggestions

- If the disk usage exceeds 85%, you are advised to expand disk capacity and attach the newly purchased disks to ECSs to associate with the cluster.
- The procedure for attaching disks and setting parameters may vary depending on the site environment.

15.2.3 Replacing a Disk in an MRS Cluster (Applicable to 2.x and Earlier)

Issue

A disk is not accessible.

Symptom

A user created an MRS cluster with local disks. A disk of a core node in this cluster is damaged, resulting in file read failures.

Cause Analysis

The disk hardware is faulty.

Procedure

NOTE

This procedure is applicable to analysis clusters earlier than MRS 3.*x*. If you need to replace disks for a streaming cluster or hybrid cluster, contact technical support.

Step 1 Log in to .

Step 2 Choose **Hosts**, click the name of the host to be decommissioned, click **RegionServer** in the **Roles** list, click **More**, and select **Decommission**.

Step 3 Choose **Hosts**, click the name of the host to be decommissioned, click **DataNode** in the **Roles** list, click **More**, and select **Decommission**.

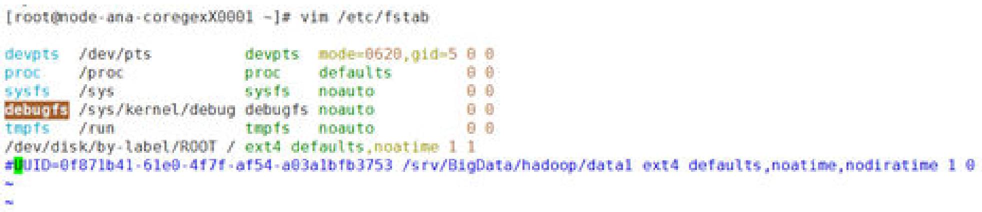
Step 4 Choose **Hosts**, click the name of the host to be decommissioned, click **NodeManager** in the **Roles** list, click **More**, and select **Decommission**.

NOTE

If this host still runs other instances, perform the similar operation to decommission the instances.

Step 5 Run the **vim /etc/fstab** command to comment out the mount point of the faulty disk.

Figure 15-1 Commenting out the mount point of the faulty disk



```
[root@node-ana-coregexX0001 ~]# vim /etc/fstab

devpts   /dev/pts          devpts  mode=0620,gid=5 0 0
proc     /proc             proc    defaults        0 0
sysfs    /sys              sysfs   noauto          0 0
debugfs  /sys/kernel/debug debugfs noauto          0 0
tmpfs    /run              tmpfs   noauto          0 0
/dev/disk/by-label/ROOT / ext4 defaults,noatime 1 1
#UUID=0f871b41-61e0-4f7f-af54-a03a1bfb3753 /srv/BigData/hadoop/data1 ext4 defaults,noatime,nodiratime 1 0
~
~
```

Step 6 Migrate the user data on the faulty disk (for example, **/srv/BigData/hadoop/data1/**).

Step 7 Log in to the MRS console.

Step 8 On the cluster details page, click the **Nodes** tab.

Step 9 Click the node whose disk is to be replaced to go to the ECS console. Click **Stop** to stop the node.

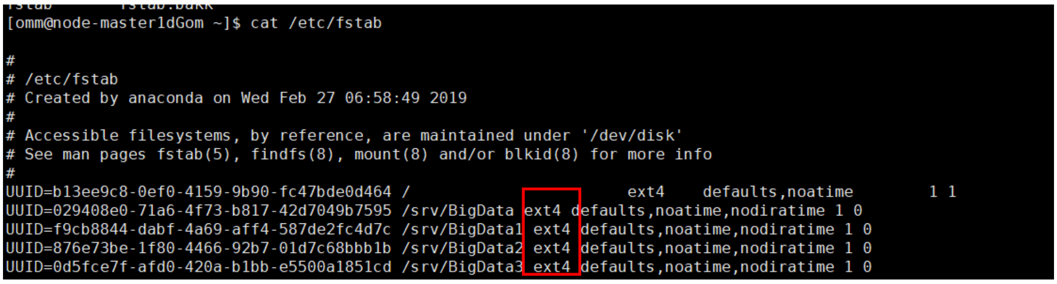
Step 10 Contact technical support to replace the disk in the background.

Step 11 On the ECS console, click **Start** to start the node where the disk has been replaced.

Step 12 Run the **fdisk -l** command to view the new disk.

Step 13 Run the **cat /etc/fstab** command to obtain the drive letter.

Figure 15-2 Obtaining the drive letter



```
fstab       fstab.bakk
[omm@node-master1dGom ~]$ cat /etc/fstab

#
# /etc/fstab
# Created by anaconda on Wed Feb 27 06:58:49 2019
#
# Accessible filesystems, by reference, are maintained under '/dev/disk'
# See man pages fstab(5), findfs(8), mount(8) and/or blkid(8) for more info
#
UUID=b13ee9c8-0ef0-4159-9b90-fc47bde0d464 /                       ext4    defaults,noatime        1 1
UUID=029408e0-71a6-4f73-b817-42d7049b7595 /srv/BigData ext4 defaults,noatime,nodiratime 1 0
UUID=f9cb8844-dabf-4a69-aff4-587de2fc4d7c /srv/BigData1 ext4 defaults,noatime,nodiratime 1 0
UUID=876e73be-1f80-4466-92b7-01d7c68bbb1b /srv/BigData2 ext4 defaults,noatime,nodiratime 1 0
UUID=0d5fce7f-afd0-420a-b1bb-e5500a1851cd /srv/BigData3 ext4 defaults,noatime,nodiratime 1 0
```

Step 14 Use the corresponding drive letter to format the new disk.

Example: **mkfs.ext4 /dev/sdh**

Step 15 Run the following command to attach the new disk.

mount *New disk Mount point*

Example: **mount /dev/sdh /srv/BigData/hadoop/data1**

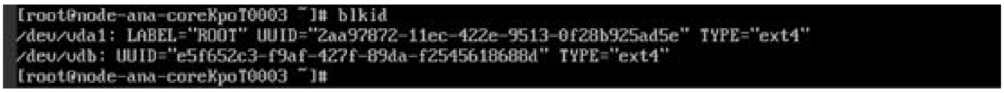
Step 16 Run the following command to grant the **omm** user permission to the new disk:

chown omm:wheel *Mount point*

Example: **chown -R omm:wheel /srv/BigData/hadoop/data1**

Step 17 Add the UUID of the new disk to the **fstab** file.

1. Run the **blkid** command to check the UUID of the new disk.

   ```
   [root@node-ana-coreKpoT0003 ~]# blkid
   /dev/vda1: LABEL="ROOT" UUID="2aa97872-11ec-422e-9513-0f28b925ad5e" TYPE="ext4"
   /dev/vdb: UUID="e5f652c3-f9af-427f-89da-f2545618688d" TYPE="ext4"
   [root@node-ana-coreKpoT0003 ~]#
   ```

2. Open the **/etc/fstab** file and add the following information:

   UUID=*New disk UUID* **/srv/BigData/hadoop/data1 ext4 defaults,noatime,nodiratime 1 0**

Step 18 (Optional) Create a log directory.

mkdir -p /srv/BigData/Bigdata

chown omm:ficommon /srv/BigData/Bigdata

chmod 770 /srv/BigData/Bigdata

NOTE

Run the following command to check whether symbolic links to **Bigdata** logs exist. If yes, skip this step.

ll /var/log

Step 19 Log in to .

Step 20 Choose **Hosts**, click the name of the host to be recommissioned, click **RegionServer** in the **Roles** list, click **More**, and select **Recommission**.

Step 21 Choose **Hosts**, click the name of the host to be recommissioned, click **DataNode** in the **Roles** list, click **More**, and select **Recommission**.

Step 22 Choose **Hosts**, click the name of the host to be recommissioned, click **NodeManager** in the **Roles** list, click **More**, and select **Recommission**.

NOTE

If this host still runs other instances, perform the similar operation to recommission the instances.

Step 23 Choose **Services** > **HDFS**. In the **HDFS Summary** area on the **Service Status** page, check whether **Missing Blocks** is **0**.

- If **Missing Blocks** is **0**, no further action is required.
- If **Missing Blocks** is not **0**, contact technical support.

----End

15.2.4 Replacing a Disk in an MRS Cluster (Applicable to 3.x)

Issue

A disk is not accessible.

Symptom

A user created an MRS cluster with local disks. A disk of a core node in this cluster is damaged, resulting in file read failures.

Cause Analysis

The disk hardware is faulty.

Procedure

NOTE

This procedure is applicable to troubleshooting disk hardware faults of core and task nodes in MRS clusters using local disks (ECSs of D, I, IR, and KI series).

Kafka does not support disk replacement. If the node that stores Kafka data is faulty, contact technical support.

Step 1 Log in to .

Step 2 Choose **Hosts** and click the name of the faulty host. In the **Instance** area, click **DataNode**. Then on the page that is displayed, click **More** and select **Decommission**.

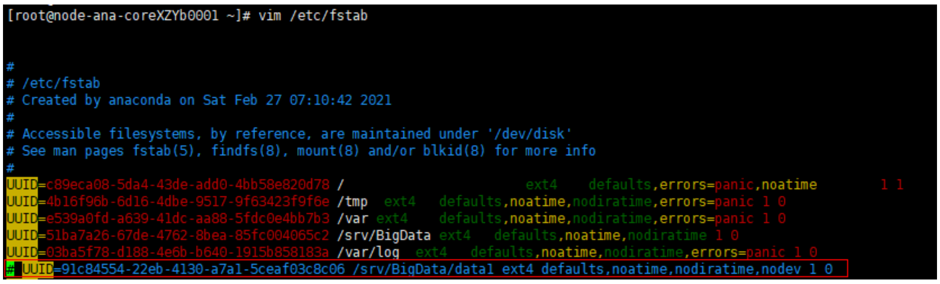
NOTE

- If this host accommodates DataNode, NodeManager, RegionServer, and ClickHouseServer instances, decommission these instances by referring to this step.
- In versions later than MRS 3.1.2, the ClickHouseServer role instance can be decommissioned.

Step 3 Choose **Hosts**, select the faulty host, click **More**, and select **Stop All Instances**.

Step 4 Run the **vim /etc/fstab** command to comment out the mount point of the faulty disk.

Figure 15-3 Commenting out the mount point of the faulty disk



```
[root@node-ana-coreXZYb0001 ~]# vim /etc/fstab


#
# /etc/fstab
# Created by anaconda on Sat Feb 27 07:10:42 2021
#
# Accessible filesystems, by reference, are maintained under '/dev/disk'
# See man pages fstab(5), findfs(8), mount(8) and/or blkid(8) for more info
#
UUID=c89eca08-5da4-43de-add0-4bb58e820d78 /                       ext4    defaults,errors=panic,noatime        1 1
UUID=4b16f96b-6d16-4dbe-9517-9f63423f9f6e /tmp  ext4   defaults,noatime,nodiratime,errors=panic 1 0
UUID=e539a0fd-a639-41dc-aa88-5fdc0e4bb7b3 /var ext4    defaults,noatime,nodiratime,errors=panic 1 0
UUID=51ba7a26-67de-4762-8bea-85fc004065c2 /srv/BigData ext4   defaults,noatime,nodiratime 1 0
UUID=03ba5f78-d188-4e6b-b640-1915b858183a /var/log  ext4   defaults,noatime,nodiratime,errors=panic 1 0
# UUID=91c84554-22eb-4130-a7a1-5ceaf03c8c06 /srv/BigData/data1 ext4 defaults,noatime,nodiratime,nodev 1 0
```

Step 5 If the old disk is still accessible, migrate user data on the old disk (for example, **/srv/BigData/data1/**).

cp -r ***Mount point of the old disk Temporary data storage directory***

Example: **cp -r /srv/BigData/data1 /tmp/**

Step 6 Log in to the MRS console.

Step 7 On the cluster details page, click the **Nodes** tab.

Step 8 Click the node whose disk is to be replaced to go to the ECS console. Click **Stop** to stop the node.

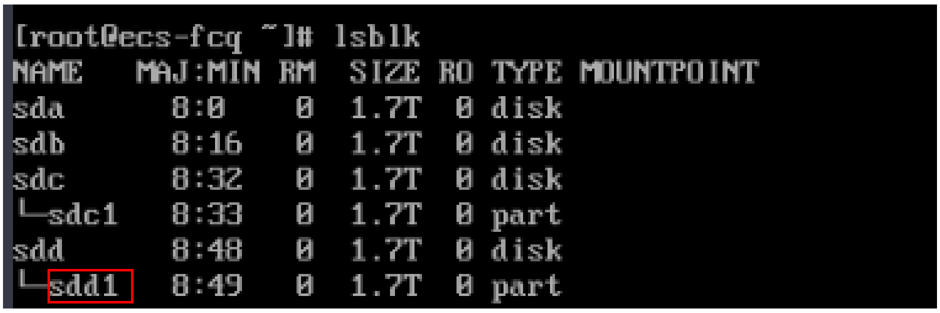
Step 9 Contact technical support to replace the disk in the background.

Step 10 On the ECS console, click **Start** to start the node where the disk has been replaced.

Step 11 Initialize the Linux data disk.

Step 12 Run the **lsblk** command to view information about the new disk partition.

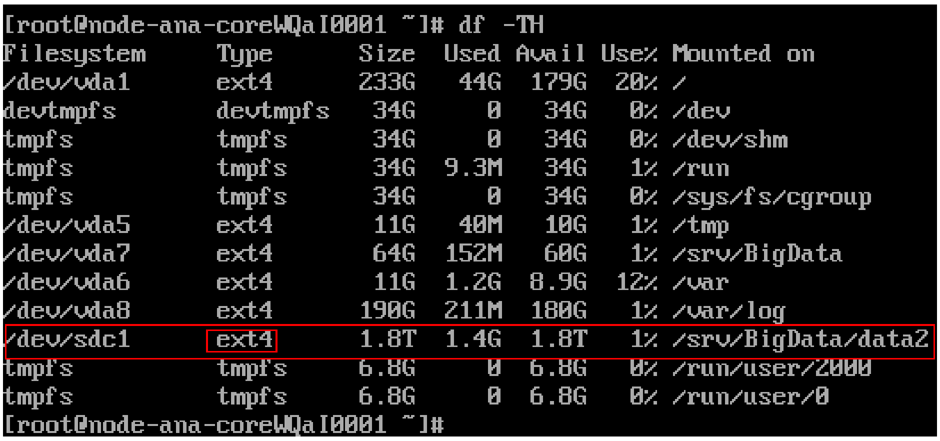
Figure 15-4 Viewing the new disk partition



```
[root@ecs-fcq ~]# lsblk
NAME    MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT
sda       8:0    0  1.7T  0 disk
sdb       8:16   0  1.7T  0 disk
sdc       8:32   0  1.7T  0 disk
└─sdc1    8:33   0  1.7T  0 part
sdd       8:48   0  1.7T  0 disk
└─sdd1    8:49   0  1.7T  0 part
```

Step 13 Run the **df -TH** command to obtain the file system type.

Figure 15-5 Obtaining the file system type



```
[root@node-ana-coreWQaI0001 ~]# df -TH
Filesystem     Type      Size  Used Avail Use% Mounted on
/dev/vda1      ext4      233G   44G  179G  20% /
devtmpfs       devtmpfs   34G     0   34G   0% /dev
tmpfs          tmpfs      34G     0   34G   0% /dev/shm
tmpfs          tmpfs      34G  9.3M   34G   1% /run
tmpfs          tmpfs      34G     0   34G   0% /sys/fs/cgroup
/dev/vda5      ext4       11G   40M   10G   1% /tmp
/dev/vda7      ext4       64G  152M   60G   1% /srv/BigData
/dev/vda6      ext4       11G  1.2G  8.9G  12% /var
/dev/vda8      ext4      190G  211M  180G   1% /var/log
/dev/sdc1      ext4      1.8T  1.4G  1.8T   1% /srv/BigData/data2
tmpfs          tmpfs     6.8G     0  6.8G   0% /run/user/2000
tmpfs          tmpfs     6.8G     0  6.8G   0% /run/user/0
[root@node-ana-coreWQaI0001 ~]#
```

Step 14 Format the new disk partition based on the obtained file system type.

Example: **mkfs.ext4 /dev/sdd1**

Step 15 Run the following command to mount the new disk:

mount *New disk Mount point*

Example: **mount /dev/sdd1 /srv/BigData/data1**

NOTE

If the disk cannot be mounted, run the following command to reload the configuration and mount it again:

systemctl daemon-reload

Step 16 Run the following command to grant the **omm** user permission to the new disk:

chown omm:wheel *Mount point*

Example: **chown -R omm:wheel /srv/BigData/data1**

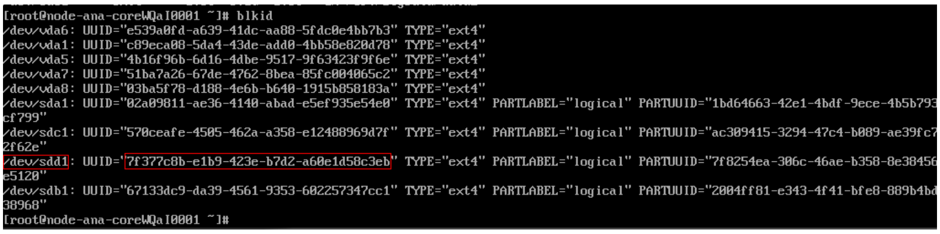
Step 17 Migrate user data from the old disk (for example, **/srv/BigData/data1/**) to the new disk.

cp -r*Temporary data storage directory Mount point of the new disk*

Example: **cp -r /tmp/data1/* /srv/BigData/data1/**

Step 18 Add the UUID of the new disk to the **fstab** file.

1. Run the **blkid** command to check the UUID of the new disk.



```
[root@node-ana-coreWQaI0001 ~]# blkid
/dev/vda6: UUID="e539a0fd-a639-41dc-aa88-5fdc0e4bb7b3" TYPE="ext4"
/dev/vda1: UUID="c89eca08-5da4-43de-add0-4bb58e820d78" TYPE="ext4"
/dev/vda5: UUID="4b16f96b-6d16-4dbe-9517-9f63423f9f6e" TYPE="ext4"
/dev/vda7: UUID="51ba7a26-67de-4762-8bea-85fc004065c2" TYPE="ext4"
/dev/vda8: UUID="03ba5f78-d188-4e6b-b640-1915b858183a" TYPE="ext4"
/dev/sda1: UUID="02a09811-ae36-4140-abad-e5ef935e54e0" TYPE="ext4" PARTLABEL="logical" PARTUUID="1bd64663-42e1-4bdf-9ece-4b5b793
cf799"
/dev/sdc1: UUID="570ceafe-4505-462a-a358-e12488969d7f" TYPE="ext4" PARTLABEL="logical" PARTUUID="ac309415-3294-47c4-b089-ae39fc7
2f62e"
/dev/sdd1: UUID="7f377c8b-e1b9-423e-b7d2-a60e1d58c3eb" TYPE="ext4" PARTLABEL="logical" PARTUUID="7f8254ea-306c-46ae-b358-8e38456
e5120"
/dev/sdb1: UUID="67133dc9-da39-4561-9353-602257347cc1" TYPE="ext4" PARTLABEL="logical" PARTUUID="2004ff81-e343-4f41-bfe8-889b4bd
38968"
[root@node-ana-coreWQaI0001 ~]#
```

2. Open the **/etc/fstab** file and add the following information:
UUID=*UUID of the new disk* /srv/BigData/data1 ext4 defaults,noatime,nodiratime,nodev 1 0

Step 19 Log in to .

Step 20 Choose **Hosts** and click the name of the host to be recommissioned. In the **Instance** area, click **DataNode**. Then on the page that is displayed, click **More** and select **Recommission**.

NOTE

- If this host accommodates DataNode, NodeManager, RegionServer, and ClickHouseServer instances, recommission these instances by referring to this step.
- In versions later than MRS 3.1.2, the ClickHouseServer role instance can be recommissioned.

Step 21 Choose **Hosts**, select the faulty host, click **More**, and select **Start All Instances**.

Step 22 Choose **Cluster** > **HDFS**. In the **Basic Information** area on the **Dashboard** page, check whether **Missing Blocks** is **0**.

- If **Missing Blocks** is **0**, no further action is required.
- If **Missing Blocks** is not **0**, contact technical support.

----End

15.2.5 MRS Backup Failure

Issue

MRS backup keeps failing.

Symptom

MRS backup keeps failing.

Cause Analysis

The backup directory is connected to the system disk using a soft link. As a result, if the system disk is full, the backup fails.

Procedure

Step 1 Check whether the backup directory is connected to the system disk using a soft disk.

1. Log in to the active and standby Master nodes in the cluster as user **root**.
2. Run the **df -h** command to check the storage usage of the system disk.
3. Run the **ll /srv/BigData/LocalBackup** command to check whether the backup directory is connected to **/opt/Bigdata/LocalBackup** using a soft link.

   Check whether the backup file is connected to the system disk using a soft link and whether the system disk has sufficient space. If the soft link is used for connecting to the system disk and the system disk space is insufficient, go to **Step 2**. If the soft link is not used, the failure is not caused by insufficient system disk space. Contact technical support for troubleshooting.

Step 2 Move historical backup data to a new directory on the data disk.

1. Log in to the Master node as user **root**.
2. Run the **su - omm** command to switch to user **omm**.
3. Run the **rm -rf /srv/BigData/LocalBackup** command to delete the soft link of the backup directory.
4. Run the **mkdir -p /srv/BigData/LocalBackup** command to create a backup directory.
5. Run the **mv /opt/Bigdata/LocalBackup/* /srv/BigData/LocalBackup/** command to move the historical backup data to the new directory.

----End

15.2.6 Inconsistency Between df and du Command Output on the Core Node

Issue

The capacity displayed in the **df** command output on the Core node is inconsistent with that displayed in the **du** command output.

Symptom

After the **df** and **du** commands are executed, the values of the Core node capacity displayed are different.

The disk usage of the **/srv/BigData/hadoop/data1/** directory queried by running the **df -h** command differs greatly from that queried by running the **du -sh /srv/BigData/hadoop/data1/** command. The difference is greater than 10 GB.

Cause Analysis

The **lsof |grep deleted** command output indicates that a large number of log files in the directory are in the deleted state.

When some Spark tasks are running for a long time, some containers in the tasks keep running and logs are continuously generated. When printing logs, the executor of Spark uses the log4j log scrolling function to output logs to the **stdout** file. The container also monitors this file. As a result, the file is monitored by two processes at the same time. When one process scrolls according to the configuration, the earliest log file is deleted, but the other process still occupies the file handle. As a result, a file in the deleted state is generated.

Procedure

Change the output directory name for executor logs of Spark.

1. Open the log configuration file. By default, the configuration file is located in *<Client address>***/Spark/spark/conf/log4j-executor.properties**.
2. Change the name of the log output file.

   For example, change **log4j.appender.sparklog.File = ${spark.yarn.app.container.log.dir}/stdout** to **log4j.appender.sparklog.File = ${spark.yarn.app.container.log.dir}/stdout.log**.
3. Save the configuration and exit.
4. Submit the tasks again.

15.2.7 Disassociating a Subnet from the ACL Network

Scenarios

You can disassociate a subnet from the ACL network when necessary.

Procedure

Step 1 Log in to the management console.

Step 2 On the console homepage, under **Network**, click **Virtual Private Cloud**.

Step 3 In the navigation tree on the left, choose **Network ACL**.

Step 4 Locate the target network ACL in the right pane, and click the network ACL name to switch to the network ACL details page.

Step 5 On the displayed page, click the **Associated Subnets** tab.

Step 6 On the **Associated Subnets** page, locate the target network ACL and click **Disassociate** in the **Operation** column.

Step 7 Click **OK**.

----End

15.2.8 MRS Becomes Abnormal After hostname Modification

Issue

What should I do if MRS becomes abnormal after **hostname** is modified?

Symptom

MRS becomes abnormal after **hostname** is modified.

Possible Cause

The **hostname** modification causes compatibility problems and faults.

Procedure

Step 1 Log in to any node in the cluster as user **root**.

Step 2 Run the **cat /etc/hosts** command on the node to check the value of **hostname** of each node and set the **newhostname** variable based on the value.

Step 3 Run the **sudo hostnamectl set-hostname ${newhostname}** command on the node where **hostname** is modified to restore the correct hostname.

NOTE

${newhostname}: new value of **hostname**

Step 4 After the modification, log in to the node where **hostname** is modified, and check whether the new hostname takes effect.

----End

15.2.9 DataNode Restarts Unexpectedly

Symptom

A DataNode is restarted unexpectedly, but no manual restart operation is performed for the DataNode.

Cause Analysis

Possible causes:

- **OOM of the Java process is killed.**

  In general, the OMM Killer is configured for Java processes to detect and kill OOM. The OOM log is printed in the out log. In this case, you can view the run log (for example, the DataNode's log path is **/var/log/Bigdata/hdfs/dn/hadoop-omm-datanode-*hostname*.log**) to check whether OutOfMemory is printed.

- **DataNode is manually killed or killed by another process.**

  Check the DataNode run log file **/var/log/Bigdata/hdfs/dn/hadoop-omm-datanode-*hostname*.log**. It is found that the health check fails after "RECEIVED SIGNAL 15" is received. In the following example, the DataNode is killed at 11:04:48 and then started at 11:06:52 two minutes later.

  ```
  2018-12-06 11:04:48,433 | ERROR | SIGTERM handler | RECEIVED SIGNAL 15: SIGTERM |
  LogAdapter.java:69
  2018-12-06 11:04:48,436 | INFO  | Thread-1 | SHUTDOWN_MSG:
  /************************************************************
  SHUTDOWN_MSG: Shutting down DataNode at 192-168-235-85/192.168.235.85
  ************************************************************/ | LogAdapter.java:45
  2018-12-06 11:06:52,744 | INFO  | main | STARTUP_MSG:
  ```

  According to the logs, DataNode was closed and then the health check reported the exception. After 2 minutes, NodeAgent started the DataNode process.

Procedure

Add the rule for recording the kill command in the audit log of the operating system. The process that delivers the kill command will be recorded in the audit log.

Operation impact

- Printing audit logs affects operating system performance. However, analysis result shows that the impact is less than 1%.
- Printing audit log occupies some disk space. The logs to be printed are within megabytes. By default, the aging mechanism and the mechanism for checking the remaining disk space are configured. Therefore, the disk space will not be used up.

Locating Method

Perform the following operations on nodes that may restart the DataNode process:

Step 1 Log in to the node as the **root** user and run the **service auditd status** command to check the service status.

```
Checking for service auditd  running
```

If the service is not started, run the **service auditd restart** command to restart the service. The command execution takes less than 1 second and has no impact on the system.

```
Shutting down auditd done
Starting auditd done
```

Step 2 The audit rule of the **kill** command is temporarily added to audit logs.

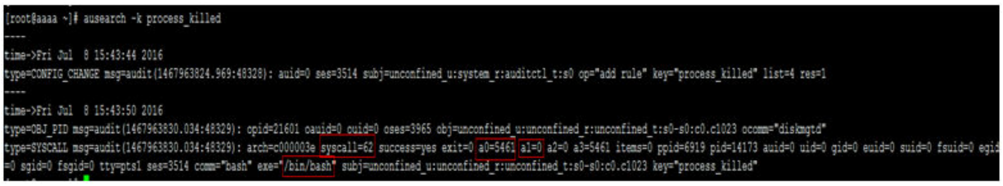
Add an audit rule:

auditctl -a exit,always -F arch=b64 -S kill -S tkill -S tgkill -F a1!=0 -k process_killed

View the rule:

auditctl -l

Step 3 If a process is killed due to an exception, you can run the **ausearch -k process_killed** command to query the kill history.



```
[root@aaaa ~]# ausearch -k process_killed
----
time->Fri Jul  8 15:43:44 2016
type=CONFIG_CHANGE msg=audit(1467963824.969:48328): auid=0 ses=3514 subj=unconfined_u:system_r:auditctl_t:s0 op="add rule" key="process_killed" list=4 res=1
----
time->Fri Jul  8 15:43:50 2016
type=OBJ_PID msg=audit(1467963830.034:48329): opid=21601 oauid=0 ouid=0 oses=3965 obj=unconfined_u:unconfined_r:unconfined_t:s0-s0:c0.c1023 ocomm="diskmgtd"
type=SYSCALL msg=audit(1467963830.034:48329): arch=c000003e syscall=62 success=yes exit=0 a0=5461 a1=0 a2=0 a3=5461 items=0 ppid=6919 pid=14173 auid=0 uid=0 gid=0 euid=0 suid=0 fsuid=0 egid
=0 sgid=0 fsgid=0 tty=pts1 ses=3514 comm="bash" exe="/bin/bash" subj=unconfined_u:unconfined_r:unconfined_t:s0-s0:c0.c1023 key="process_killed"
```

NOTE

a0 is the PID (hexadecimal) of the process that is killed, and **a1** is the semaphore of the kill command.

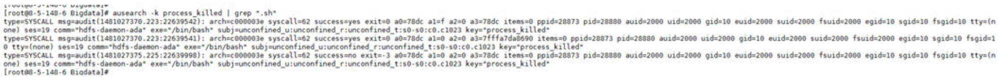
----End

Verification

Step 1 Restart an instance of the node on MRS Manager, for example, DataNode.

Step 2 Run the **ausearch -k process_killed** command to check whether logs are printed.

The following is an example of the **ausearch -k process_killed |grep ".sh"** command. The command output indicates that the **hdfs-daemon-ada*** script closed the DataNode process.



```
[root@8-5-148-6 Bigdata]# ausearch -k process_killed | grep ".sh"
type=SYSCALL msg=audit(1481027370.223:22639542): arch=c000003e syscall=62 success=yes exit=0 a0=78dc a1=f a2=0 a3=78dc items=0 ppid=28873 pid=28880 auid=2000 uid=2000 gid=10 euid=2000 suid=2000 fsuid=2000 egid=10 sgid=10 fsgid=10 tty=(n
one) ses=19 comm="hdfs-daemon-ada" exe="/bin/bash" subj=unconfined_u:unconfined_r:unconfined_t:s0-s0:c0.c1023 key="process_killed"
type=SYSCALL msg=audit(1481027370.223:22639541): arch=c000003e syscall=62 success=yes exit=0 a0=78dc a1=0 a2=0 a3=7fffa7da8690 items=0 ppid=28873 pid=28880 auid=2000 uid=2000 gid=10 euid=2000 suid=2000 fsuid=2000 egid=10 sgid=10 fsgid=1
0 tty=(none) ses=19 comm="hdfs-daemon-ada" exe="/bin/bash" subj=unconfined_u:unconfined_r:unconfined_t:s0-s0:c0.c1023 key="process_killed"
type=SYSCALL msg=audit(1481027375.225:22639998): arch=c000003e syscall=62 success=no exit=-3 a0=78dc a1=0 a2=0 a3=78dc items=0 ppid=28873 pid=28880 auid=2000 uid=2000 gid=10 euid=2000 suid=2000 fsuid=2000 egid=10 sgid=10 fsgid=10 tty=(n
one) ses=19 comm="hdfs-daemon-ada" exe="/bin/bash" subj=unconfined_u:unconfined_r:unconfined_t:s0-s0:c0.c1023 key="process_killed"
[root@8-5-148-6 Bigdata]#
```

----End

Stop auditing the **kill** command.

Step 1 Run the **service auditd restart** command. The temporarily added kill command audit logs are cleared automatically.

Step 2 Run the **auditctl -l** command. If no information about killing a process is returned, the rule is cleared successfully.

----End

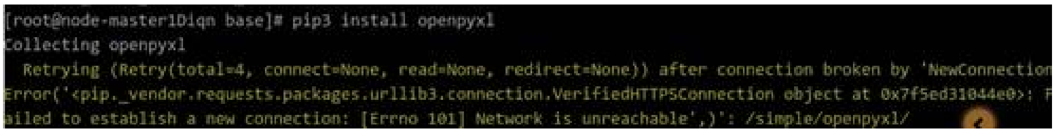
15.2.10 Network Is Unreachable When Using pip3 to Install the Python Package in an MRS Cluster

Issue

When the Python package is installed using pip3, an error message is displayed, indicating that the network is unreachable.

Symptom

When a user runs the pip3 install command to install the Python package, an error message is displayed, indicating that the network is unreachable. For details, see the following figure:



```
[root@node-master1Diqn base]# pip3 install openpyxl
Collecting openpyxl
  Retrying (Retry(total=4, connect=None, read=None, redirect=None)) after connection broken by 'NewConnection
Error('<pip._vendor.requests.packages.urllib3.connection.VerifiedHTTPSConnection object at 0x7f5ed31044e0>: F
ailed to establish a new connection: [Errno 101] Network is unreachable',)': /simple/openpyxl/
```

Cause Analysis

The customer does not bind an EIP to the Master node.

Procedure

Step 1 Log in to the MRS management console.

Step 2 Choose **Clusters** > **Active Clusters**, select the faulty cluster, and click its name to check the **Basic Information** on the **Dashboard** tab page.

Step 3 On the **Nodes** tab page, click the name of a Master node in the Master node group to log in to the ECS management console.

Step 4 Click the **EIPs** tab and click **Bind EIP** to bind an EIP to the ECS.

Step 5 Log in to the Master node and run the **pip3 install** command to install the Python package.

----End

15.2.11 Failed to Download the MRS Cluster Client

Issue

On the local Master host, a user attempts to download an MRS cluster client for another remote host. However, the system displays a message indicating that the network or parameter is abnormal.

Symptom

On the local Master host, a user attempts to download an MRS cluster client for another remote host. However, the system displays a message indicating that the network or parameter is abnormal.

Cause Analysis

- The two hosts are in different VPCs.
- The password is incorrect.
- The firewall is enabled on the remote host.

Procedure

- The two hosts are in different VPCs.

  Enable port 22 of the remote host.
- The password is incorrect.

  Check whether the password is correct. The password cannot contain special characters.
- The firewall is enabled on the remote host.

  Download the MRS cluster client to the server and run the **scp** command provided by Linux to remotely send the client to the remote host.

15.2.12 Failed to Scale Out an MRS Cluster

Issue

The MRS console is accessible and functions properly, but the MRS cluster fails to be scaled out.

Symptom

The MRS console is normal, and no alarm or error message is displayed on MRS Manager. However, an error message is displayed during cluster scale-out, indicating that the MRS cluster contains nodes that are not running.

Cause Analysis

An MRS cluster can be scaled in or out only when it is running properly. According to the error message, the possible cause is that the cluster status in the database is abnormal or is not updated. As a result, the nodes in the cluster are not in the running state.

Procedure

Step 1 Log in to the MRS console and click the cluster name to go to the cluster details page. Check that the cluster is in the **Running** state.

Step 2 Click **Nodes** to view the status of all nodes. Ensure that all nodes are in the **Running** state.

Step 3 Log in to the podMaster node in the cluster, switch to the MRS deployer node, and view the **api-gateway.log** file.

1. Run the **kubectl get pod -n mrs** command to view the **pod** of the MRS deployer node.
2. Run the **kubectl exec -ti ${***Pod of the deployer node***} -n mrs /bin/bash** command to log in to the pod. For example, run the **kubectl exec -ti mrsdeployer-78bc8c76cf-mn9ss -n mrs /bin/bash** command to access the deployer container of MRS.
3. In the **/opt/cloud/logs/apigateway** directory, view the latest **api-gateway.log** file and search for the required keyword (such as **ERROR**, **scaling**, **clusterScaling**, **HostState**, **state-check**, or the cluster ID) in the file to check the error type.
4. Rectify the fault based on the error information and perform the scale-out again.
   - If the scale-out is successful, no further action is required.
   - If the scale-out fails, go to **Step 4**.

Step 4 Run the **/opt/cloud/mysql -u${***Username***} -P${***Port***} -h${***Address***} -p${***Password***}** command to log in to the database.

Step 5 Run the **select cluster_state from cluster_detail where cluster_id=***Cluster ID***";** command to check the value of **cluster_state**.

- If the value of **cluster_state** is **2**, the cluster status is normal. Go to **Step 6**.
- If the value of **cluster_state** is not **2**, the cluster status in the database is abnormal. You can run the **update cluster_detail set cluster_state=2 where cluster_id="***Cluster ID***";** command to update the cluster status and then check the value of **cluster_state**.
  - If the value of **cluster_state** is **2**, the cluster status is normal. Go to **Step 6**.
  - If the value of **cluster_state** is not **2**, contact technical support.

Step 6 Run the **select host_status from host where cluster_di="***Cluster ID***";** command to query the cluster host status.

- If the host is in the started state, no further action is required.
- If the host is not in the started state, run the **update host set host_status='started' where cluster_id="***Cluster ID***";** command to update the host status to the database.

- If the host is in the started state, no further action is required.
- If the host is not in the started state, contact technical support.

----End

15.2.13 Error Occurs When MRS Executes the Insert Command Using Beeline

Issue

An error occurs when MRS executes the insert command using Beeline.

Symptom

When the **insert into** statement is executed in Beeline of Hive, the following error is reported:

```
Mapping run in Tez on Hive transactional table fails when data volume is high with error:
"org.apache.hadoop.hive.ql.lockmgr.LockException Reason: Transaction... already aborted, Hive SQL state
[42000]."
```

Cause Analysis

This problem is caused by improper cluster configuration and Tez resource setting.

Procedure

This problem can be solved by setting configuration parameters on Beeline.

Step 1 Set the following properties to optimize performance (you are advised to change them at the cluster level):

- Set **hive.auto.convert.sortmerge.join** to **true**.
- Set **hive.optimize.bucketmapjoin** to **true**.
- Set **hive.optimize.bucketmapjoin.sortedmerge** to **true**.

Step 2 Modify the following content to adjust the resources of Tez:

- Set **hive.tez.container.size** to the size of the Yarn container.
- Set **hive.tez.container.size** to the Yarn container size **yarn.scheduler.minimum-allocation-mb** or a smaller value (for example, a half or quarter of the Yarn container size). Ensure that the value does not exceed **yarn.scheduler.maximum-allocation-mb**.

----End

15.2.14 How Do I Upgrade EulerOS to Fix Vulnerabilities in an MRS Cluster?

Issue

EulerOS has vulnerabilities at the underlying layer. This section describes how to upgrade the OS to fix vulnerabilities for an MRS cluster.

Symptom

When the NSFOCUS software is used to test the cluster, vulnerabilities are found at the underlying layer in the EulerOS.

Cause Analysis

When the NSFOCUS software is used to test the cluster, it is found that vulnerabilities exist at the underlying layer in the EulerOS. The MRS service is deployed in the EulerOS. Therefore, the system needs to be upgraded to fix the vulnerabilities.

Procedure

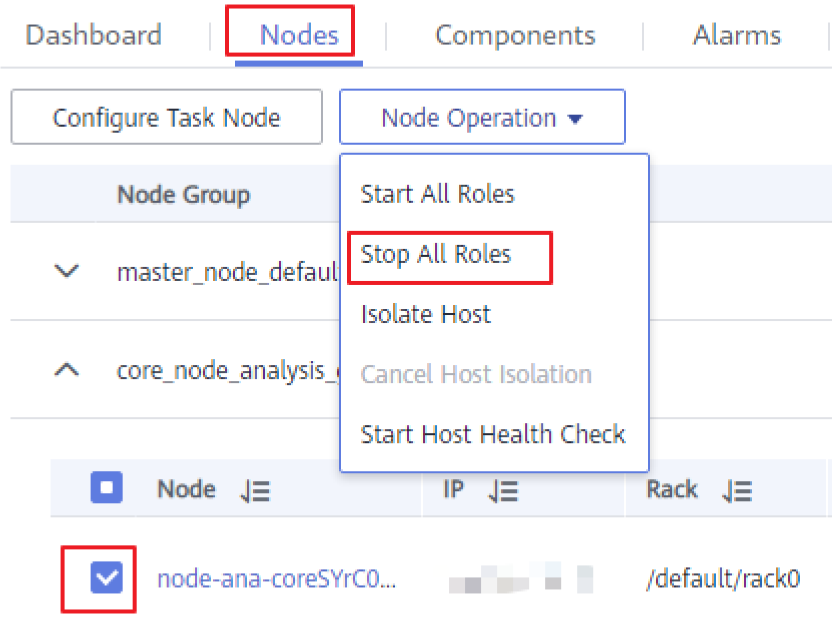
NOTE

Before fixing the vulnerability, check whether Host Security Service (HSS) is enabled. If yes, disable HSS from monitoring the MRS cluster. After the vulnerability is fixed, enable HSS again.

Step 1 Log in to the MRS console.

Step 2 Click the cluster name. On the cluster details page, click the **Nodes** tab.

Step 3 In the core node group, select a core node, click **Node Operation**, and select **Stop All Roles**.



Step 4 Remotely log in to the core node and configure the yum repository.

Step 5 Run the **uname –r** or **rpm –qa |grep kernel** command to query and record the kernel version of the current node.

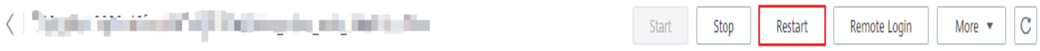
Step 6 Run the **yum update –y --skip-broken --setopt=protected_multilib=false** command to update the patch.

Step 7 After the update is complete, query the kernel version and run the **rpm -e** *Old kernel version* command to delete the old kernel version.

Step 8 On the cluster details page, click the **Nodes** tab.

Step 9 In the core node group, click the name of the core node whose patch has been updated. The ECS console is displayed.

Step 10 In the upper right corner of the page, click **Restart** to restart the core node.



Step 11 On the **Nodes** tab of the cluster details page, select the core node, click **Node Operation**, and select **Start All Roles**.

Step 12 Repeat **Step 1** to **Step 11** to upgrade other core nodes.

Step 13 After all core nodes are upgraded, upgrade the standby master node and then the active master node. For details, see **Step 1** to **Step 11**.

----End

15.2.15 Using CDM to Migrate Data to HDFS

Issue

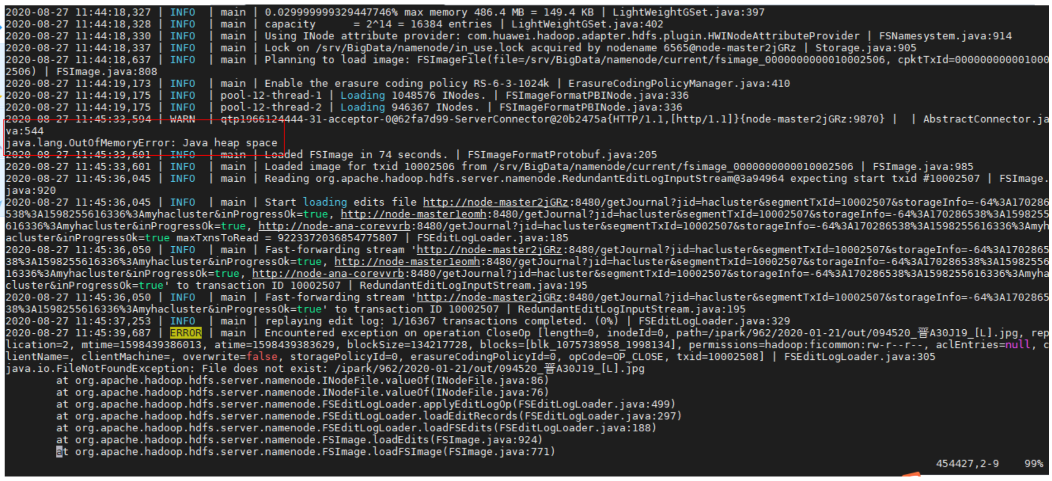
A user failed to use CDM to migrate data from an old cluster to HDFS of a new cluster.

Symptom

When CDM is used to import data from the source HDFS to the destination HDFS, the destination MRS cluster is faulty and the NameNode cannot be started.

The logs show that the **Java heap space** error is reported during the startup. The JVM parameter of the NameNode needs to be modified.

Figure 15-6 Fault logs



```
2020-08-27 11:44:18,327 | INFO  | main | 0.029999999329447746% max memory 486.4 MB = 149.4 KB | LightWeightGSet.java:397
2020-08-27 11:44:18,328 | INFO  | main | capacity      = 2^14 = 16384 entries | LightWeightGSet.java:402
2020-08-27 11:44:18,330 | INFO  | main | Using INode attribute provider: com.huawei.hadoop.adapter.hdfs.plugin.HWINodeAttributeProvider | FSNamesystem.java:914
2020-08-27 11:44:18,337 | INFO  | main | Lock on /srv/BigData/namenode/in_use.lock acquired by nodename 6565@node-master2jGRz | Storage.java:905
2020-08-27 11:44:18,637 | INFO  | main | Planning to load image: FSImageFile(file=/srv/BigData/namenode/current/fsimage_0000000000010002506, cpktTxId=000000000001000
2506) | FSImage.java:808
2020-08-27 11:44:19,173 | INFO  | main | Enable the erasure coding policy RS-6-3-1024k | ErasureCodingPolicyManager.java:410
2020-08-27 11:44:19,175 | INFO  | pool-12-thread-1 | Loading 1048576 INodes. | FSImageFormatPBINode.java:336
2020-08-27 11:44:19,175 | INFO  | pool-12-thread-2 | Loading 946367 INodes. | FSImageFormatPBINode.java:336
2020-08-27 11:45:33,594 | WARN  | qtp1966124444-31-acceptor-0@62fa7d99-ServerConnector@20b2475a{HTTP/1.1,[http/1.1]}{node-master2jGRz:9870} |  | AbstractConnector.ja
va:544
java.lang.OutOfMemoryError: Java heap space
2020-08-27 11:45:33,601 | INFO  | main | Loaded FSImage in 74 seconds. | FSImageFormatProtobuf.java:205
2020-08-27 11:45:33,601 | INFO  | main | Loaded image for txid 10002506 from /srv/BigData/namenode/current/fsimage_0000000000010002506 | FSImage.java:985
2020-08-27 11:45:36,045 | INFO  | main | Reading org.apache.hadoop.hdfs.server.namenode.RedundantEditLogInputStream@3a94964 expecting start txid #10002507 | FSImage.
java:920
2020-08-27 11:45:36,045 | INFO  | main | Start loading edits file http://node-master2jGRz:8480/getJournal?jid=hacluster&segmentTxId=10002507&storageInfo=-64%3A170286
538%3A1598255616336%3Amyhacluster&inProgressOk=true, http://node-master1eomh:8480/getJournal?jid=hacluster&segmentTxId=10002507&storageInfo=-64%3A170286538%3A1598255
616336%3Amyhacluster&inProgressOk=true, http://node-ana-corevvrb:8480/getJournal?jid=hacluster&segmentTxId=10002507&storageInfo=-64%3A170286538%3A1598255616336%3Amyh
acluster&inProgressOk=true maxTxnsToRead = 9223372036854775807 | FSEditLogLoader.java:185
2020-08-27 11:45:36,050 | INFO  | main | Fast-forwarding stream 'http://node-master2jGRz:8480/getJournal?jid=hacluster&segmentTxId=10002507&storageInfo=-64%3A1702865
38%3A1598255616336%3Amyhacluster&inProgressOk=true, http://node-master1eomh:8480/getJournal?jid=hacluster&segmentTxId=10002507&storageInfo=-64%3A170286538%3A15982556
16336%3Amyhacluster&inProgressOk=true, http://node-ana-corevvrb:8480/getJournal?jid=hacluster&segmentTxId=10002507&storageInfo=-64%3A170286538%3A1598255616336%3Amyha
cluster&inProgressOk=true' to transaction ID 10002507 | RedundantEditLogInputStream.java:195
2020-08-27 11:45:36,050 | INFO  | main | Fast-forwarding stream 'http://node-master2jGRz:8480/getJournal?jid=hacluster&segmentTxId=10002507&storageInfo=-64%3A1702865
38%3A1598255616336%3Amyhacluster&inProgressOk=true' to transaction ID 10002507 | RedundantEditLogInputStream.java:195
2020-08-27 11:45:37,253 | INFO  | main | replaying edit log: 1/16367 transactions completed. (0%) | FSEditLogLoader.java:329
2020-08-27 11:45:39,687 | ERROR | main | Encountered exception on operation CloseOp [length=0, inodeId=0, path=/ipark/962/2020-01-21/out/094520_晋A30J19_[L].jpg, rep
lication=2, mtime=1598439386013, atime=1598439383629, blockSize=134217728, blocks=[blk_1075738958_1998134], permissions=hadoop:ficommon:rw-r--r--, aclEntries=null, c
lientName=, clientMachine=, overwrite=false, storagePolicyId=0, erasureCodingPolicyId=0, opCode=OP_CLOSE, txid=10002508] | FSEditLogLoader.java:305
java.io.FileNotFoundException: File does not exist: /ipark/962/2020-01-21/out/094520_晋A30J19_[L].jpg
        at org.apache.hadoop.hdfs.server.namenode.INodeFile.valueOf(INodeFile.java:86)
        at org.apache.hadoop.hdfs.server.namenode.INodeFile.valueOf(INodeFile.java:76)
        at org.apache.hadoop.hdfs.server.namenode.FSEditLogLoader.applyEditLogOp(FSEditLogLoader.java:499)
        at org.apache.hadoop.hdfs.server.namenode.FSEditLogLoader.loadEditRecords(FSEditLogLoader.java:297)
        at org.apache.hadoop.hdfs.server.namenode.FSEditLogLoader.loadFSEdits(FSEditLogLoader.java:188)
        at org.apache.hadoop.hdfs.server.namenode.FSImage.loadEdits(FSImage.java:924)
        at org.apache.hadoop.hdfs.server.namenode.FSImage.loadFSImage(FSImage.java:771)
                                                                                                                                454427,2-9    99%
```

Cause Analysis

When the user uses CDM to migrate data, the HDFS data volume is too large. As a result, a stack exception occurs when metadata is merged.

Procedure

Step 1 Go to the HDFS service configuration page.

- For versions earlier than MRS 2.0.1: Log in to MRS Manager, choose **Services** > **HDFS** > **Service Configuration**, and select **All** from the **Basic** drop-down list.
- For MRS 2.0.1 or later: Click the cluster name on the MRS console, choose **Components** > **HDFS** > **Service Configuration**, and select **All** from the **Basic** drop-down list.
- For MRS 3.*x* or later: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the target cluster and choose **Services** > **HDFS** > **Configurations** > **All Configurations**.

Step 2 Search for the **GC_OPTS** parameter in **HDFS->NameNode** and increase the values of **-Xms512M** and **-Xmx512M** based on service requirements.

Step 3 Save the configuration and restart the affected services or instances.

----End

15.2.16 Alarms Are Frequently Generated in the MRS Cluster

Issue

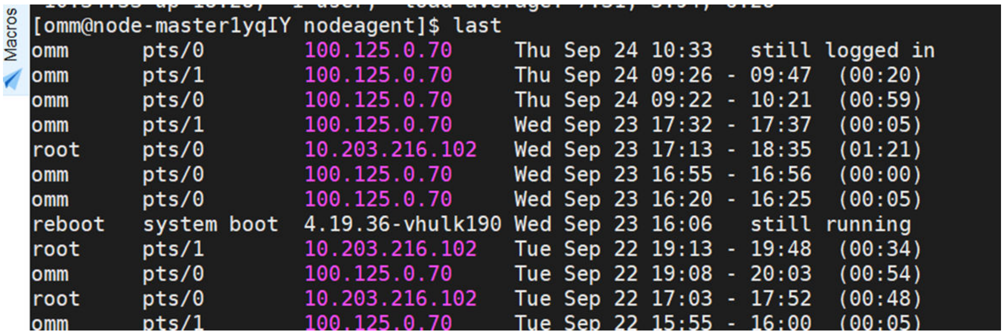
The cluster frequently reports alarms indicating that the heartbeat between the active and standby Manager nodes is interrupted, the heartbeat between the active and standby DBService nodes is interrupted, and the node is faulty. As a result, Hive is occasionally unavailable.

Symptom

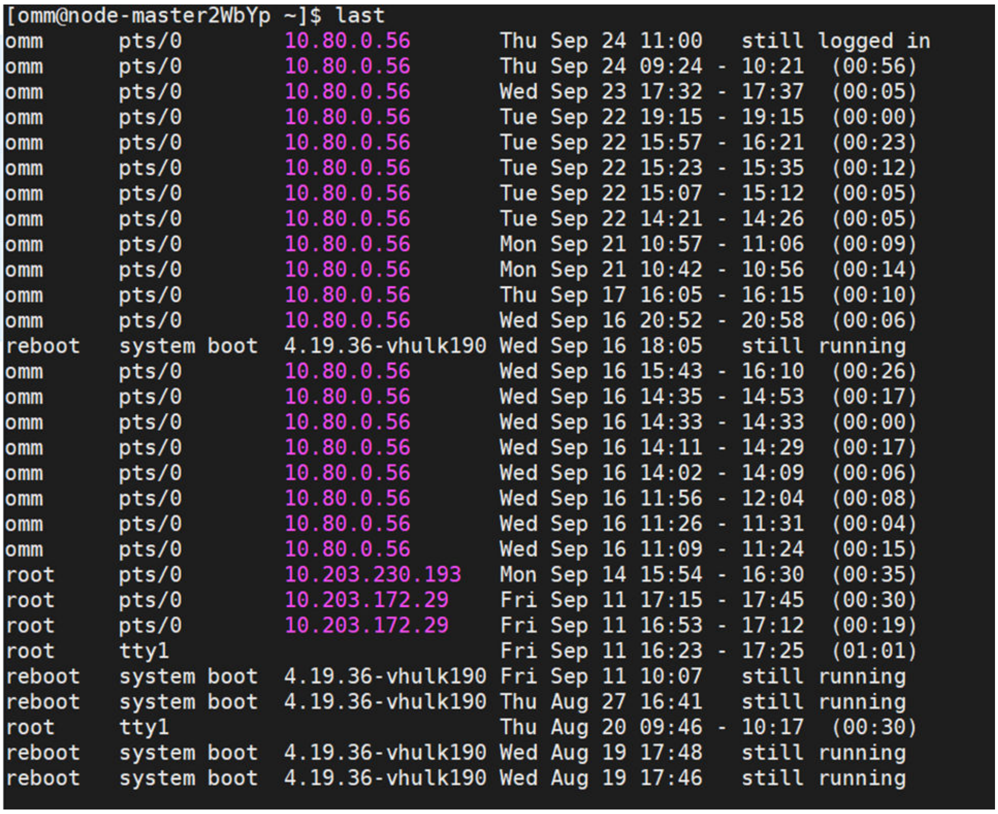
The cluster frequently reports alarms indicating that the heartbeat between the active and standby Manager nodes is interrupted, the heartbeat between the active and standby DBService nodes is interrupted, and the node is faulty. As a result, Hive is occasionally unavailable, affecting customer services

Cause Analysis

1. When the alarm is generated, the VM is restarted. The alarm is generated because the VM is restarted.

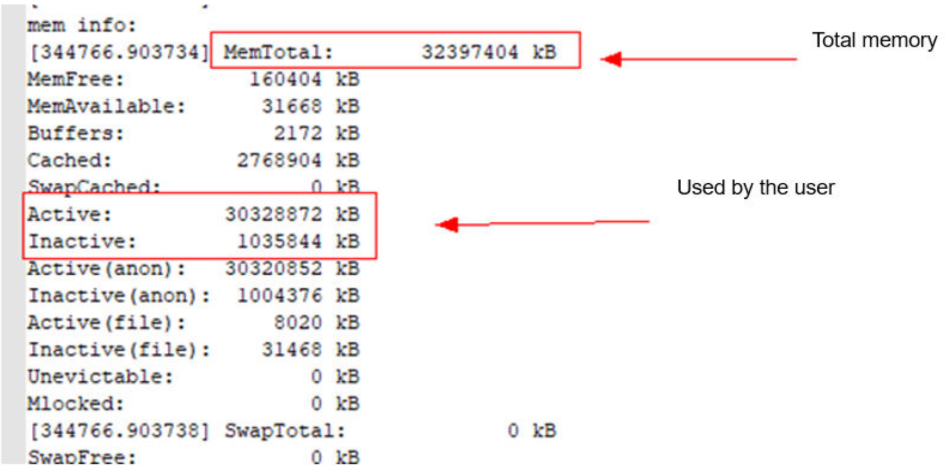


```
[omm@node-master1yqIY nodeagent]$ last
omm      pts/0        100.125.0.70     Thu Sep 24 10:33   still logged in
omm      pts/1        100.125.0.70     Thu Sep 24 09:26 - 09:47  (00:20)
omm      pts/0        100.125.0.70     Thu Sep 24 09:22 - 10:21  (00:59)
omm      pts/1        100.125.0.70     Wed Sep 23 17:32 - 17:37  (00:05)
root     pts/0        10.203.216.102   Wed Sep 23 17:13 - 18:35  (01:21)
omm      pts/0        100.125.0.70     Wed Sep 23 16:55 - 16:56  (00:00)
omm      pts/0        100.125.0.70     Wed Sep 23 16:20 - 16:25  (00:05)
reboot   system boot  4.19.36-vhulk190 Wed Sep 23 16:06   still running
root     pts/1        10.203.216.102   Tue Sep 22 19:13 - 19:48  (00:34)
omm      pts/0        100.125.0.70     Tue Sep 22 19:08 - 20:03  (00:54)
root     pts/0        10.203.216.102   Tue Sep 22 17:03 - 17:52  (00:48)
omm      pts/1        100.125.0.70     Tue Sep 22 15:55 - 16:00  (00:05)
```

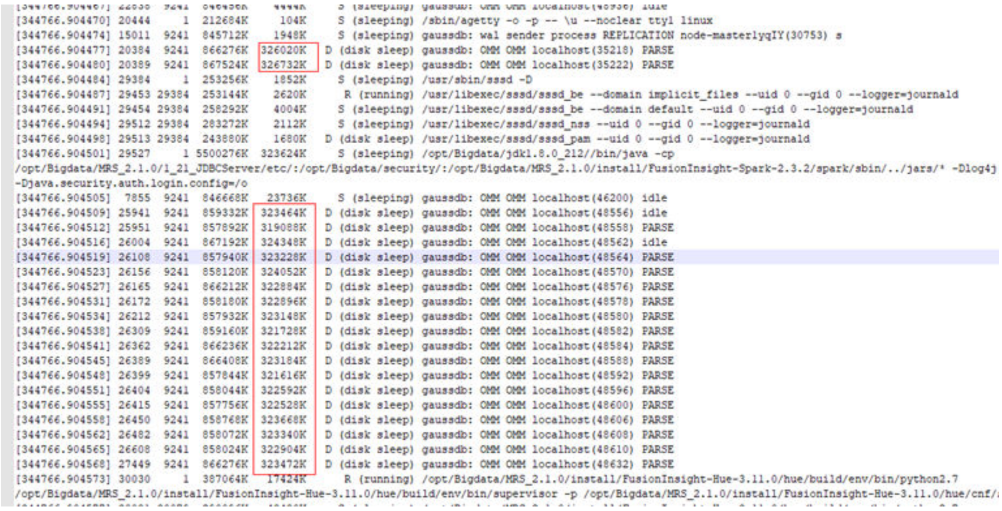


```
[omm@node-master2WbYp ~]$ last
omm      pts/0        10.80.0.56       Thu Sep 24 11:00   still logged in
omm      pts/0        10.80.0.56       Thu Sep 24 09:24 - 10:21  (00:56)
omm      pts/0        10.80.0.56       Wed Sep 23 17:32 - 17:37  (00:05)
omm      pts/0        10.80.0.56       Tue Sep 22 19:15 - 19:15  (00:00)
omm      pts/0        10.80.0.56       Tue Sep 22 15:57 - 16:21  (00:23)
omm      pts/0        10.80.0.56       Tue Sep 22 15:23 - 15:35  (00:12)
omm      pts/0        10.80.0.56       Tue Sep 22 15:07 - 15:12  (00:05)
omm      pts/0        10.80.0.56       Tue Sep 22 14:21 - 14:26  (00:05)
omm      pts/0        10.80.0.56       Mon Sep 21 10:57 - 11:06  (00:09)
omm      pts/0        10.80.0.56       Mon Sep 21 10:42 - 10:56  (00:14)
omm      pts/0        10.80.0.56       Thu Sep 17 16:05 - 16:15  (00:10)
omm      pts/0        10.80.0.56       Wed Sep 16 20:52 - 20:58  (00:06)
reboot   system boot  4.19.36-vhulk190 Wed Sep 16 18:05   still running
omm      pts/0        10.80.0.56       Wed Sep 16 15:43 - 16:10  (00:26)
omm      pts/0        10.80.0.56       Wed Sep 16 14:35 - 14:53  (00:17)
omm      pts/0        10.80.0.56       Wed Sep 16 14:33 - 14:33  (00:00)
omm      pts/0        10.80.0.56       Wed Sep 16 14:11 - 14:29  (00:17)
omm      pts/0        10.80.0.56       Wed Sep 16 14:02 - 14:09  (00:06)
omm      pts/0        10.80.0.56       Wed Sep 16 11:56 - 12:04  (00:08)
omm      pts/0        10.80.0.56       Wed Sep 16 11:26 - 11:31  (00:04)
omm      pts/0        10.80.0.56       Wed Sep 16 11:09 - 11:24  (00:15)
root     pts/0        10.203.230.193   Mon Sep 14 15:54 - 16:30  (00:35)
root     pts/0        10.203.172.29    Fri Sep 11 17:15 - 17:45  (00:30)
root     pts/0        10.203.172.29    Fri Sep 11 16:53 - 17:12  (00:19)
root     tty1                          Fri Sep 11 16:23 - 17:25  (01:01)
reboot   system boot  4.19.36-vhulk190 Fri Sep 11 10:07   still running
reboot   system boot  4.19.36-vhulk190 Thu Aug 27 16:41   still running
root     tty1                          Thu Aug 20 09:46 - 10:17  (00:30)
reboot   system boot  4.19.36-vhulk190 Wed Aug 19 17:48   still running
reboot   system boot  4.19.36-vhulk190 Wed Aug 19 17:46   still running
```

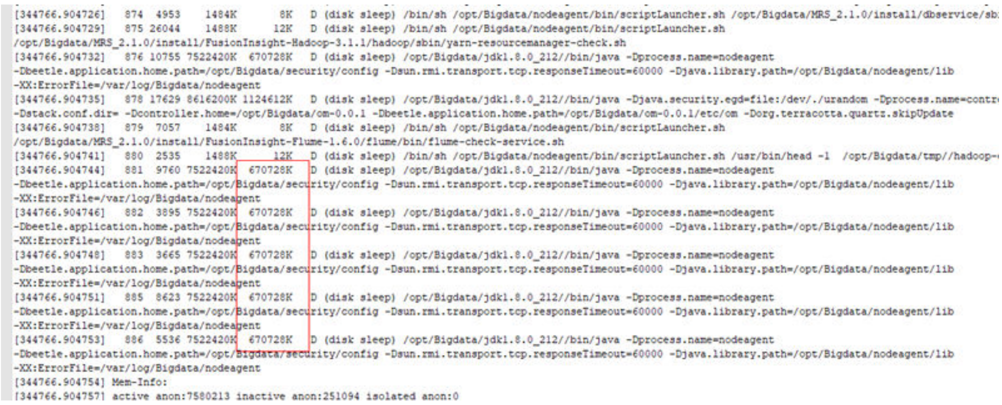
2. According to the OS analysis, the cause of the VM restart is that the node does not have available memory. Memory overflow triggers oom-killer. When the process is invoked, the process enters the **disk sleep** state. As a result, the VM restarts.



```
mem info:
[344766.903734] MemTotal:        32397404 kB
MemFree:           160404 kB
MemAvailable:       31668 kB
Buffers:             2172 kB
Cached:           2768904 kB
SwapCached:             0 kB
Active:          30328872 kB
Inactive:         1035844 kB
Active(anon):    30320852 kB
Inactive(anon):   1004376 kB
Active(file):        8020 kB
Inactive(file):     31468 kB
Unevictable:            0 kB
Mlocked:                0 kB
[344766.903738] SwapTotal:              0 kB
SwapFree:               0 kB
```



```
[344766.904470] 20444     1  212684K    104K     S (sleeping) /sbin/agetty -o -p -- \u --noclear tty1 linux
[344766.904474] 15011  9241  845712K   1948K     S (sleeping) gaussdb: wal sender process REPLICATION node-master1yqIY(30753) s
[344766.904477] 20384  9241  866276K 326020K   D (disk sleep) gaussdb: OMM OMM localhost(35218) PARSE
[344766.904480] 20389  9241  867524K 326732K   D (disk sleep) gaussdb: OMM OMM localhost(35222) PARSE
[344766.904484] 29384     1  253256K   1852K     S (sleeping) /usr/sbin/sssd -D
[344766.904487] 29453 29384  253144K   2620K      R (running) /usr/libexec/sssd/sssd_be --domain implicit_files --uid 0 --gid 0 --logger=journald
[344766.904491] 29454 29384  258292K   4004K     S (sleeping) /usr/libexec/sssd/sssd_be --domain default --uid 0 --gid 0 --logger=journald
[344766.904494] 29512 29384  283272K   2112K     S (sleeping) /usr/libexec/sssd/sssd_nss --uid 0 --gid 0 --logger=journald
[344766.904498] 29513 29384  243880K   1680K   D (disk sleep) /usr/libexec/sssd/sssd_pam --uid 0 --gid 0 --logger=journald
[344766.904501] 29527     1 5500276K 323624K     S (sleeping) /opt/Bigdata/jdk1.8.0_212//bin/java -cp
/opt/Bigdata/MRS_2.1.0/1_21_JDBCServer/etc/:/opt/Bigdata/security/:/opt/Bigdata/MRS_2.1.0/install/FusionInsight-Spark-2.3.2/spark/sbin/../jars/* -Dlog4j
-Djava.security.auth.login.config=/o
[344766.904505]  7855  9241  846668K  23736K     S (sleeping) gaussdb: OMM OMM localhost(46200) idle
[344766.904509] 25941  9241  859332K 323464K   D (disk sleep) gaussdb: OMM OMM localhost(48556) idle
[344766.904512] 25951  9241  857892K 319088K   D (disk sleep) gaussdb: OMM OMM localhost(48558) PARSE
[344766.904516] 26004  9241  867192K 324348K   D (disk sleep) gaussdb: OMM OMM localhost(48562) idle
[344766.904519] 26108  9241  857940K 323228K   D (disk sleep) gaussdb: OMM OMM localhost(48564) PARSE
[344766.904523] 26156  9241  858120K 324052K   D (disk sleep) gaussdb: OMM OMM localhost(48570) PARSE
[344766.904527] 26165  9241  866212K 322884K   D (disk sleep) gaussdb: OMM OMM localhost(48576) PARSE
[344766.904531] 26172  9241  858180K 322896K   D (disk sleep) gaussdb: OMM OMM localhost(48578) PARSE
[344766.904534] 26212  9241  857932K 323148K   D (disk sleep) gaussdb: OMM OMM localhost(48580) PARSE
[344766.904538] 26309  9241  859160K 321728K   D (disk sleep) gaussdb: OMM OMM localhost(48582) PARSE
[344766.904541] 26362  9241  866236K 322212K   D (disk sleep) gaussdb: OMM OMM localhost(48584) PARSE
[344766.904545] 26389  9241  866408K 323184K   D (disk sleep) gaussdb: OMM OMM localhost(48588) PARSE
[344766.904548] 26399  9241  857844K 321616K   D (disk sleep) gaussdb: OMM OMM localhost(48592) PARSE
[344766.904551] 26404  9241  858044K 322592K   D (disk sleep) gaussdb: OMM OMM localhost(48596) PARSE
[344766.904555] 26415  9241  857756K 322528K   D (disk sleep) gaussdb: OMM OMM localhost(48600) PARSE
[344766.904558] 26450  9241  858768K 323668K   D (disk sleep) gaussdb: OMM OMM localhost(48606) PARSE
[344766.904562] 26482  9241  858072K 323340K   D (disk sleep) gaussdb: OMM OMM localhost(48608) PARSE
[344766.904565] 26608  9241  858024K 322904K   D (disk sleep) gaussdb: OMM OMM localhost(48610) PARSE
[344766.904568] 27449  9241  866276K 323472K   D (disk sleep) gaussdb: OMM OMM localhost(48632) PARSE
[344766.904573] 30030     1  387064K  17424K      R (running) /opt/Bigdata/MRS_2.1.0/install/FusionInsight-Hue-3.11.0/hue/build/env/bin/python2.7
/opt/Bigdata/MRS_2.1.0/install/FusionInsight-Hue-3.11.0/hue/build/env/bin/supervisor -p /opt/Bigdata/MRS_2.1.0/install/FusionInsight-Hue-3.11.0/hue/cnf/
```



```
[344766.904726]   874  4953    1484K      8K   D (disk sleep) /bin/sh /opt/Bigdata/nodeagent/bin/scriptLauncher.sh /opt/Bigdata/MRS_2.1.0/install/dbservice/sb
[344766.904729]   875 26044    1488K     12K   D (disk sleep) /bin/sh /opt/Bigdata/nodeagent/bin/scriptLauncher.sh
/opt/Bigdata/MRS_2.1.0/install/FusionInsight-Hadoop-3.1.1/hadoop/sbin/yarn-resourcemanager-check.sh
[344766.904732]   876 10755 7522420K  670728K   D (disk sleep) /opt/Bigdata/jdk1.8.0_212//bin/java -Dprocess.name=nodeagent
-Dbeetle.application.home.path=/opt/Bigdata/security/config -Dsun.rmi.transport.tcp.responseTimeout=60000 -Djava.library.path=/opt/Bigdata/nodeagent/lib
-XX:ErrorFile=/var/log/Bigdata/nodeagent
[344766.904735]   878 17629 8616200K 1124612K   D (disk sleep) /opt/Bigdata/jdk1.8.0_212//bin/java -Djava.security.egd=file:/dev/./urandom -Dprocess.name=contr
-Dstack.conf.dir= -Dcontroller.home=/opt/Bigdata/om-0.0.1 -Dbeetle.application.home.path=/opt/Bigdata/om-0.0.1/etc/om -Dorg.terracotta.quartz.skipUpdate
[344766.904738]   879  7057    1484K      8K   D (disk sleep) /bin/sh /opt/Bigdata/nodeagent/bin/scriptLauncher.sh
/opt/Bigdata/MRS_2.1.0/install/FusionInsight-Flume-1.6.0/flume/bin/flume-check-service.sh
[344766.904741]   880  2535    1488K     12K   D (disk sleep) /bin/sh /opt/Bigdata/nodeagent/bin/scriptLauncher.sh /usr/bin/head -1  /opt/Bigdata/tmp//hadoop-
[344766.904744]   881  9760 7522420K  670728K   D (disk sleep) /opt/Bigdata/jdk1.8.0_212//bin/java -Dprocess.name=nodeagent
-Dbeetle.application.home.path=/opt/Bigdata/security/config -Dsun.rmi.transport.tcp.responseTimeout=60000 -Djava.library.path=/opt/Bigdata/nodeagent/lib
-XX:ErrorFile=/var/log/Bigdata/nodeagent
[344766.904746]   882  3895 7522420K  670728K   D (disk sleep) /opt/Bigdata/jdk1.8.0_212//bin/java -Dprocess.name=nodeagent
-Dbeetle.application.home.path=/opt/Bigdata/security/config -Dsun.rmi.transport.tcp.responseTimeout=60000 -Djava.library.path=/opt/Bigdata/nodeagent/lib
-XX:ErrorFile=/var/log/Bigdata/nodeagent
[344766.904748]   883  3665 7522420K  670728K   D (disk sleep) /opt/Bigdata/jdk1.8.0_212//bin/java -Dprocess.name=nodeagent
-Dbeetle.application.home.path=/opt/Bigdata/security/config -Dsun.rmi.transport.tcp.responseTimeout=60000 -Djava.library.path=/opt/Bigdata/nodeagent/lib
-XX:ErrorFile=/var/log/Bigdata/nodeagent
[344766.904751]   885  8623 7522420K  670728K   D (disk sleep) /opt/Bigdata/jdk1.8.0_212//bin/java -Dprocess.name=nodeagent
-Dbeetle.application.home.path=/opt/Bigdata/security/config -Dsun.rmi.transport.tcp.responseTimeout=60000 -Djava.library.path=/opt/Bigdata/nodeagent/lib
-XX:ErrorFile=/var/log/Bigdata/nodeagent
[344766.904753]   886  5536 7522420K  670728K   D (disk sleep) /opt/Bigdata/jdk1.8.0_212//bin/java -Dprocess.name=nodeagent
-Dbeetle.application.home.path=/opt/Bigdata/security/config -Dsun.rmi.transport.tcp.responseTimeout=60000 -Djava.library.path=/opt/Bigdata/nodeagent/lib
-XX:ErrorFile=/var/log/Bigdata/nodeagent
[344766.904754] Mem-Info:
[344766.904757] active_anon:7580213 inactive_anon:251094 isolated_anon:0
```

3. Check the processes that occupy the memory. It is found that the processes that occupy the memory are normal service processes.

Conclusion: The VM memory cannot meet service requirements.

Procedure

- You are advised to expand the node memory.
- You are advised to disable unnecessary services to avoid this problem.

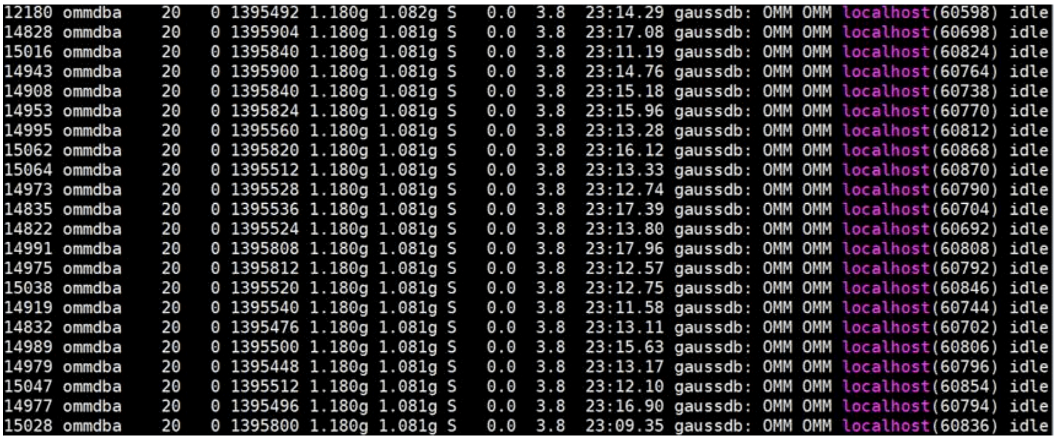
15.2.17 Memory Usage of the PMS Process Is High

Issue

What can I do if the memory usage of the active Master node is high?

Symptom

The memory usage of the active Master node is high. The **top -c** command output shows that the following idle processes occupy a large amount of memory:



```
12180 ommdba    20   0 1395492 1.180g 1.082g S   0.0  3.8  23:14.29 gaussdb: OMM OMM localhost(60598) idle
14828 ommdba    20   0 1395904 1.180g 1.081g S   0.0  3.8  23:17.08 gaussdb: OMM OMM localhost(60698) idle
15016 ommdba    20   0 1395840 1.180g 1.081g S   0.0  3.8  23:11.19 gaussdb: OMM OMM localhost(60824) idle
14943 ommdba    20   0 1395900 1.180g 1.081g S   0.0  3.8  23:14.76 gaussdb: OMM OMM localhost(60764) idle
14908 ommdba    20   0 1395840 1.180g 1.081g S   0.0  3.8  23:15.18 gaussdb: OMM OMM localhost(60738) idle
14953 ommdba    20   0 1395824 1.180g 1.081g S   0.0  3.8  23:15.96 gaussdb: OMM OMM localhost(60770) idle
14995 ommdba    20   0 1395560 1.180g 1.081g S   0.0  3.8  23:13.28 gaussdb: OMM OMM localhost(60812) idle
15062 ommdba    20   0 1395820 1.180g 1.081g S   0.0  3.8  23:16.12 gaussdb: OMM OMM localhost(60868) idle
15064 ommdba    20   0 1395512 1.180g 1.081g S   0.0  3.8  23:13.33 gaussdb: OMM OMM localhost(60870) idle
14973 ommdba    20   0 1395528 1.180g 1.081g S   0.0  3.8  23:12.74 gaussdb: OMM OMM localhost(60790) idle
14835 ommdba    20   0 1395536 1.180g 1.081g S   0.0  3.8  23:17.39 gaussdb: OMM OMM localhost(60704) idle
14822 ommdba    20   0 1395524 1.180g 1.081g S   0.0  3.8  23:13.80 gaussdb: OMM OMM localhost(60692) idle
14991 ommdba    20   0 1395808 1.180g 1.081g S   0.0  3.8  23:17.96 gaussdb: OMM OMM localhost(60808) idle
14975 ommdba    20   0 1395812 1.180g 1.081g S   0.0  3.8  23:12.57 gaussdb: OMM OMM localhost(60792) idle
15038 ommdba    20   0 1395520 1.180g 1.081g S   0.0  3.8  23:12.75 gaussdb: OMM OMM localhost(60846) idle
14919 ommdba    20   0 1395540 1.180g 1.081g S   0.0  3.8  23:11.58 gaussdb: OMM OMM localhost(60744) idle
14832 ommdba    20   0 1395476 1.180g 1.081g S   0.0  3.8  23:13.11 gaussdb: OMM OMM localhost(60702) idle
14989 ommdba    20   0 1395500 1.180g 1.081g S   0.0  3.8  23:15.63 gaussdb: OMM OMM localhost(60806) idle
14979 ommdba    20   0 1395448 1.180g 1.081g S   0.0  3.8  23:13.17 gaussdb: OMM OMM localhost(60796) idle
15047 ommdba    20   0 1395512 1.180g 1.081g S   0.0  3.8  23:12.10 gaussdb: OMM OMM localhost(60854) idle
14977 ommdba    20   0 1395496 1.180g 1.081g S   0.0  3.8  23:16.90 gaussdb: OMM OMM localhost(60794) idle
15028 ommdba    20   0 1395800 1.180g 1.081g S   0.0  3.8  23:09.35 gaussdb: OMM OMM localhost(60836) idle
```

Cause Analysis

- PostgreSQL cache: In addition to common execution plan cache and data cache, PostgreSQL provides cache mechanisms such as **catalog** and **relation** to improve the efficiency of generating execution plans. In the persistent connection scenario, some of the caches are not released. As a result, the persistent connection may occupy a large amount of memory.
- PMS is a monitoring process of MRS. This process frequently creates table partitions or new tables. The PostgreSQL caches the metadata of the objects accessed by the current session, and the connections in the database connection pool of the PMS exist for a long time. Therefore, the memory occupied by the connections gradually increases.

Procedure

Step 1 Log in to the active Master node as user **root**.

Step 2 Run the following command to query the PMS process ID:

ps -ef | grep =pmsd |grep -v grep

Step 3 Run the following command to stop the PMS process. In the command, **PID** indicates the PMS process ID obtained in **Step 2**.

kill -9 PID

Step 4 Wait for the PMS process to automatically start.

It takes 2 to 3 minutes to start PMS. PMS is a monitoring process. Restarting PMS does not affect big data services.

----End

15.2.18 High Memory Usage of the Knox Process

Issue

The memory usage of the knox process is high.

Symptom

The memory usage of the active Master node is high. The **top -c** command output shows that the memory usage of the knox process exceeds 4 GB.

Cause Analysis

The memory is not separately configured for the knox process. The process automatically allocates available memory based on the system memory size. As a result, the knox process occupies a large amount of memory.

Procedure

Step 1 Log in to the Master nodes as user **root**.

Step 2 Open the **/opt/knox/bin/gateway.sh** file. Search for **APP_MEM_OPTS**, and set its value to **-Xms3072m -Xmx4096m**.

Step 3 Log in to Manager and click **Hosts**. Find the IP address of the active Master node (that is, the node with a solid star before the hostname), and log in to the background of the node.

Step 4 Run the following commands to restart the process:

su - omm

sh /opt/knox/bin/restart-knox.sh

----End

15.2.19 It Takes a Long Time to Access HBase from a Client Installed on a Node Outside the Security Cluster

Issue

The cluster client is installed on a node outside the security cluster. When a user runs the **hbase shell** command on the client to access HBase, it is found that the access is very slow.

Symptom

A user creates a security cluster, installs a cluster client on a node outside the cluster, and runs the **hbase shell** command to access HBase. It is found that the access to HBase is very slow.

Cause Analysis

Kerberos authentication is required for a security cluster. You need to configure the **hosts** file on the client node to ensure that the access speed is not affected. An example of the **hosts** configuration is as follows:

```
1.1.1.1 hadoop.782670e3_1364_47e2_8c70_1b61bb80479c.com
1.1.1.1 hadoop.hadoop.com
1.1.1.1 hacluster
1.1.1.1 haclusterX
1.1.1.1 haclusterX1
```

```
1.1.1.1 haclusterX2
1.1.1.1 haclusterX3
1.1.1.1 haclusterX4
1.1.1.1 ClusterX
1.1.1.1 manager
ip1 hostname1
ip2 hostname2
ip3 hostname3
ip4 hostname4
```

Procedure

Copy the content of the **hosts** file on the cluster node to the **hosts** file on the node where the client is installed.

15.2.20 How Do I Locate a Job Submission Failure?

Symptom

A user cannot submit jobs through DGC or on the MRS console.

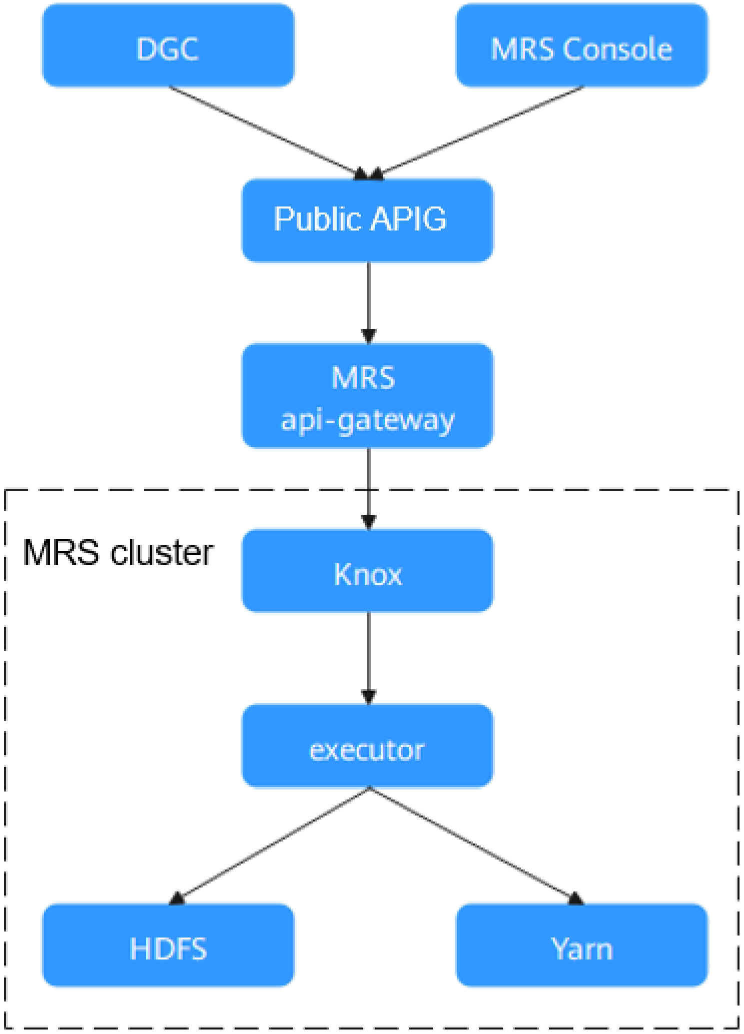
Impact

Jobs cannot be submitted, and services are interrupted.

Introduction to the Operation Process

1. All requests pass through APIG gateway and are restricted by the flow control configured on APIG.
2. APIG forwards the request to the api-gateway of the MRS management plane.
3. The API node on the MRS management plane polls the Knox of the active and standby OMS nodes to determine the Knox of the active OMS node.
4. MRS management-plane API submits a task to Knox of the active OMS.
5. Knox forwards requests to the Executor process on the current node.
6. The executor process submits a task to Yarn.

Figure 15-7 Job process



Procedure

Make preparations:

- Check whether the job is submitted through DGC or on the MRS console.
- Prepare the information listed in **Table 15-1**.

Table 15-1 Items to be prepared before the rectification

No.	Projects	Operation Mode
1	Cluster account information	Apply for **password** of user **admin** in the cluster.
2	Node account information	Apply for the passwords of users **omm** and **root** of cluster nodes.

No.	Projects	Operation Mode
3	Secure Shell (SSH) remote login tool	Prepare such tools as PuTTY or SecureCRT.
4	Client	Install the client.

Step 1 Locate the cause of the exception.

View the error code received in the job log and check whether the error code belongs to APIG or MRS.

- If the error code is a public APIG error code (starting with "APIGW"), contact public APIG maintenance personnel.
- If an error occurs on MRS, go to the next step.

Step 2 Check the running status of services and processes.

1. Log in to Manager and check whether a service fault occurs. If a job-related service fault or an underlying basic service fault occurs, rectify the fault.
2. Check whether a critical alarm is generated.
3. Log in to the active Master node.
4. Run the following command to check whether the OMS status is normal and whether the executor and knox processes on the active OMS node are normal: The knox is in active-active mode, and the executor is in single-active mode.

   /opt/Bigdata/om-0.0.1/sbin/status-oms.sh
5. Run the **jmap -heap** *PID* command as user **omm** to check the memory usage of the Knox and Executor processes. If the old-generation memory usage is 99.9%, the memory overflow occurs.

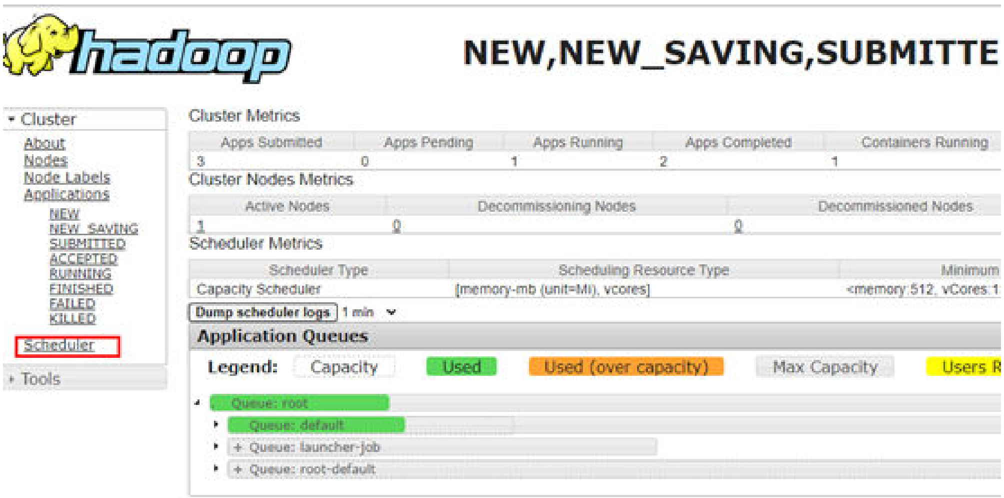
   Run the **netstat -anp | grep 8181 | grep LISTEN** command to query the PID of the executor process.

   Run the **ps -ef|grep knox | grep -v grep** command to query the PID of the knox process.

   If the memory overflows, run the **jmap -dump:format=b,file=/home/omm/temp.bin** *PID* command to export the memory information and restart the process.
6. View the native Yarn page to check the queue resource usage and whether the task is submitted to Yarn.

   On the native Yarn page: choose **Components** > **Yarn** > **ResourceManager WebUI** > **ResourceManager (Active)**.

Figure 15-8 Queue resource usage on the Yarn page



Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running
3	0	1	2	1

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
1	0	0

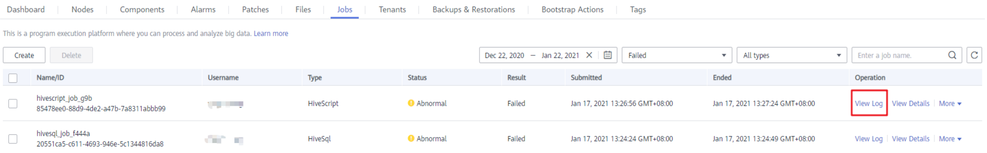
Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:512, vCores:1

Step 3 Locate the fault causing the task submission failure.

1. Log in to the MRS management console and click the cluster name to go to the cluster details page.
2. On the **Jobs** tab page, locate the row that contains the target job and click **View Log** in the **Operation** column.

Figure 15-9 View the logs



This is a program execution platform where you can process and analyze big data. Learn more

Name/ID	Username	Type	Status	Result	Submitted	Ended	Operation
hivescript_job_g9b 85478ee0-88d9-4de2-a47b-7a8311abbb99	[redacted]	HiveScript	Abnormal	Failed	Jan 17, 2021 13:26:56 GMT+08:00	Jan 17, 2021 13:27:24 GMT+08:00	View Log | View Details | More
hivesql_job_f444a 20551ca5-c611-4693-946e-5c1344816da8	[redacted]	HiveSql	Abnormal	Failed	Jan 17, 2021 13:24:24 GMT+08:00	Jan 17, 2021 13:24:49 GMT+08:00	View Log | View Details | More

3. If there is no log or the log information is not detailed, copy the job ID in the **Name/ID** column.
4. Run the following command on the active OMS node to check whether the task request is sent to the KNOX. If the request is not sent to the KNOX, the KNOX may be faulty. In this case, restart the KNOX to rectify the fault.

   grep "mrsjob" /var/log/Bigdata/knox/logs/gateway-audit.log | tail -10
5. Search for the job ID in the Executor log and view the error information.

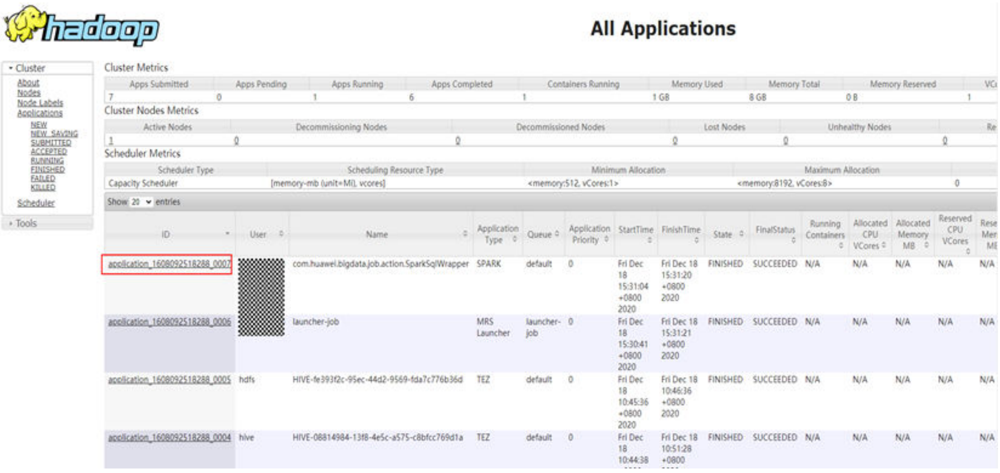
   Log file path: **/var/log/Bigdata/executor/logs/exe.log**
6. Modify the **/opt/executor/webapps/executor/WEB-INF/classes/log4j.properties** file to enable the debug log of the executor. Submit the test task and view the executor log. Confirm the error reported during job submission.

   Log file path: **/var/log/Bigdata/executor/logs/exe.log**
7. If an error occurs in the executor, run the following command to print the jstack information of the executor and check the current execution status of the thread:

   jstack PID > xxx.log
8. On the cluster details page, click the **Jobs** tab. Locate the row that contains the target job, and click **View Details** in the **Operation** column to obtain the actual job ID (**applicationID**).

9. On the cluster details page, choose **Components** > **Yarn** > **ResourceManager WebUI** > **ResourceManager (Active)**. On the native Yarn page that is displayed, click **applicationID**.

Figure 15-10 Yarn applications



All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VC
7	0	1	6	1	1 GB	8 GB	0 B	1

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Re
1	0	0	0	0	0

Scheduler Metrics

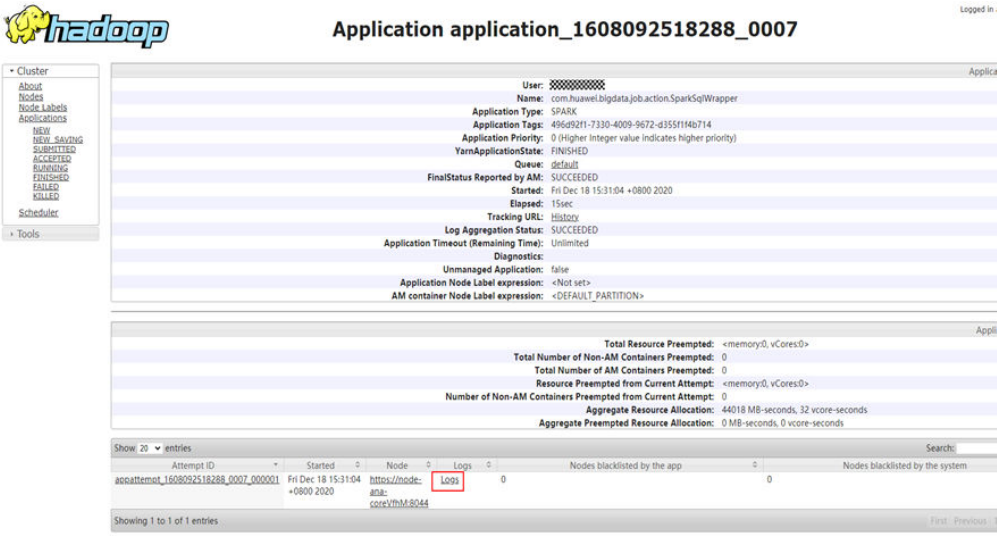
Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:512, vCores:1>	<memory:8192, vCores:8>	0

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores	Rese Mem MB
application_1608092518288_0007	[redacted]	com.huawei.bigdata.job.action.SparkSqlWrapper	SPARK	default	0	Fri Dec 18 15:31:04 +0800 2020	Fri Dec 18 15:31:20 +0800 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A
application_1608092518288_0006	[redacted]	launcher-job	MRS Launcher	launcher-job	0	Fri Dec 18 15:30:41 +0800 2020	Fri Dec 18 15:31:21 +0800 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A
application_1608092518288_0005	hdfs	HIVE-fe393f2c-95ec-44d2-9569-fda7c776b36d	TEZ	default	0	Fri Dec 18 10:45:36 +0800 2020	Fri Dec 18 10:46:36 +0800 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A
application_1608092518288_0004	hive	HIVE-08814984-13f8-4e5c-a575-c8bfcc769d1a	TEZ	default	0	Fri Dec 18 10:44:38	Fri Dec 18 10:51:28 +0800	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A

10. View logs on the task details page.

Figure 15-11 Task logs



Application application_1608092518288_0007

User: [redacted]
Name: com.huawei.bigdata.job.action.SparkSqlWrapper
Application Type: SPARK
Application Tags: 496d92f1-7330-4009-9672-d355f1f4b714
Application Priority: 0 (Higher Integer value indicates higher priority)
YarnApplicationState: FINISHED
Queue: default
FinalStatus Reported by AM: SUCCEEDED
Started: Fri Dec 18 15:31:04 +0800 2020
Elapsed: 15sec
Tracking URL: History
Log Aggregation Status: SUCCEEDED
Application Timeout (Remaining Time): Unlimited
Diagnostics:
Unmanaged Application: false
Application Node Label expression: <Not set>
AM container Node Label expression: <DEFAULT_PARTITION>

Total Resource Preempted: <memory:0, vCores:0>
Total Number of Non-AM Containers Preempted: 0
Total Number of AM Containers Preempted: 0
Resource Preempted from Current Attempt: <memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt: 0
Aggregate Resource Allocation: 44018 MB-seconds, 32 vcore-seconds
Aggregate Preempted Resource Allocation: 0 MB-seconds, 0 vcore-seconds

Show 20 entries Search:

Attempt ID	Started	Node	Logs	Nodes blacklisted by the app	Nodes blacklisted by the system
appattempt_1608092518288_0007_000001	Fri Dec 18 15:31:04 +0800 2020	https://node-ana-coreVfhM:8044	Logs	0	0

Showing 1 to 1 of 1 entries First Previous

----End

15.2.21 OS Disk Space Is Insufficient Due to Oversized HBase Log Files

Issue

The space of the **/var/log** partition on the system disk is insufficient.

Symptom

The **/var/log/Bigdata/hbase/*/hbase-omm-*.out** log file is too large, causing insufficient space of the **/var/log** partition on the system disk.

Cause Analysis

During the long-term running of HBase, the OS periodically deletes the **/tmp/.java_pid*** files created by the JVM. The HBase memory monitoring uses the **jinfo** command, which depends on the **/tmp/.java_pid*** file. If the file does not exist, the **jinfo** command runs **kill -3** to print the jstack information to the **.out** log file. As a result, the **.out** log file becomes oversize as time goes by.

Procedure

On each node hosting the HBase instance, deploy a scheduled task to periodically clear the **.out** log file. For example, log in to the HBase instance node and run the **crontab -e** command to add a scheduled task to clear the **.out** log file at 00:00:00 every day.

crontab -e

00 00 * * * for file in `ls /var/log/Bigdata/hbase/*/hbase-omm-*.out`; do echo "" > $file; done

NOTE

If large **.out** files are generated frequently, you can clear the files multiple times every day or adjust the automatic clearing policy of the OS.

15.2.22 Failed to Delete a New Tenant on FusionInsight Manager

Symptom

A user fails to delete a tenant created on the **Tenant Resources** page of FusionInsight Manager, and an error message is displayed.

Cause Analysis

When a tenant is created, its role is generated. The role will be deleted first when the tenant is deleted. If the component that supports permission configuration is abnormal, the resource permission of the role fails to be deleted.

Procedure

Step 1 Log in to FusionInsight Manager and choose **System** > **Permission** > **Role**.

Step 2 Click **Create Role**. In the **Configure Resource Permission** area, click the cluster name to check the components available for resource permission configuration.

Step 3 Choose **Cluster** > **Services** and check that the running status of these components is **Normal**.

Step 4 (Optional) If the running status is not **Normal**, start or repair the component until its running status becomes **Normal**.

Step 5 Delete the tenant again.

----End

15.3 Using Alluixo

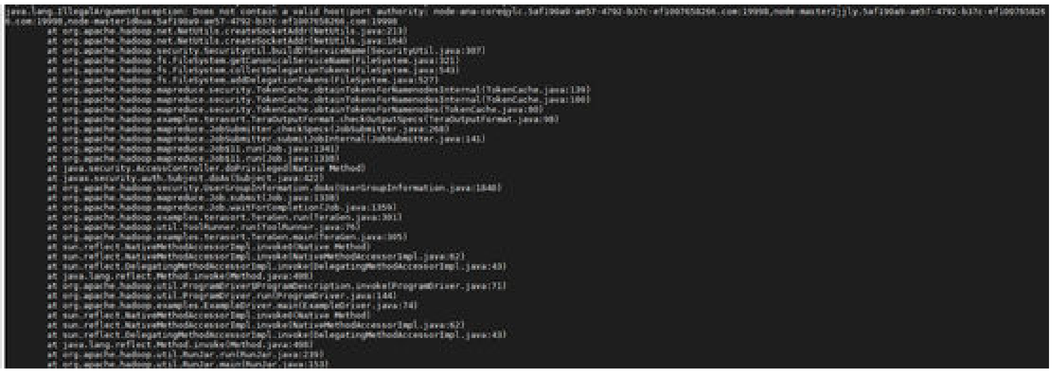
15.3.1 Error Message "Does not contain a valid host:port authority" Is Reported When Alluixo Is in HA Mode

Issue

Error message "Does not contain a valid host:port authority" is reported for Alluixo in HA mode in a security cluster.

Symptom

Error message "Does not contain a valid host:port authority" is reported for Alluixo in HA mode in a security cluster.



```
java.lang.IllegalArgumentException: Does not contain a valid host:port authority: node-ana-coreglc.5af190a9-ae57-4792-b37c-ef1007658266.com:19998,node-master2jjly.5af190a9-ae57-4792-b37c-ef100765826
6.com:19998,node-master1Bua.5af190a9-ae57-4792-b37c-ef1007658266.com:19998
        at org.apache.hadoop.net.NetUtils.createSocketAddr(NetUtils.java:213)
        at org.apache.hadoop.net.NetUtils.createSocketAddr(NetUtils.java:164)
        at org.apache.hadoop.security.SecurityUtil.buildDTServiceName(SecurityUtil.java:307)
        at org.apache.hadoop.fs.FileSystem.getCanonicalServiceName(FileSystem.java:321)
        at org.apache.hadoop.fs.FileSystem.collectDelegationTokens(FileSystem.java:543)
        at org.apache.hadoop.fs.FileSystem.addDelegationTokens(FileSystem.java:527)
        at org.apache.hadoop.mapreduce.security.TokenCache.obtainTokensForNamenodesInternal(TokenCache.java:139)
        at org.apache.hadoop.mapreduce.security.TokenCache.obtainTokensForNamenodesInternal(TokenCache.java:100)
        at org.apache.hadoop.mapreduce.security.TokenCache.obtainTokensForNamenodes(TokenCache.java:80)
        at org.apache.hadoop.examples.terasort.TeraOutputFormat.checkOutputSpecs(TeraOutputFormat.java:98)
        at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:268)
        at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:141)
        at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1341)
        at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1338)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1840)
        at org.apache.hadoop.mapreduce.Job.submit(Job.java:1338)
        at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1359)
        at org.apache.hadoop.examples.terasort.TeraGen.run(TeraGen.java:301)
        at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)
        at org.apache.hadoop.examples.terasort.TeraGen.main(TeraGen.java:305)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.hadoop.util.ProgramDriver$ProgramDescription.invoke(ProgramDriver.java:71)
        at org.apache.hadoop.util.ProgramDriver.run(ProgramDriver.java:144)
        at org.apache.hadoop.examples.ExampleDriver.main(ExampleDriver.java:74)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.hadoop.util.RunJar.run(RunJar.java:239)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:153)
```

Cause Analysis

org.apache.hadoop.security.SecurityUtil.buildDTServiceName does not support multiple alluxiomaster addresses in the URI.

Procedure

Use **alluxio:///** or **alluxio://<***IP address or hostname of the active AlluxioMaster***>: 19998/** for access.

15.4 Using ClickHouse

15.4.1 ClickHouse Fails to Start Due to Incorrect Data in ZooKeeper

Symptom

An instance node in the ClickHouse cluster fails to start. The startup log of the instance node contains error information similar to the following:

```
2021.03.15 21:01:19.816593 [ 11111 ] {} <Error> Application: DB::Exception:
The local set of parts of table DEFAULT.lineorder doesn't look like the set ofdoesn't look like the set of
parts in ZooKeeper: 59.99 million rows of 59.99 million total rows in
filesystem are suspicious. There are 30 unexpected parts with 59986052 rows
(14 of them is not just-written with 59986052 rows), 0 missing parts (with 0
blocks).: Cannot attach table `DEFAULT`.`lineorder` from metadata file
...
: while loading database
```

Cause Analysis

When a ClickHouse instance is abnormal, the ReplicatedMergeTree engine table is repeatedly created in the cluster, and then deleted. The creation and deletion of the ReplicatedMergeTree engine table causes data error in ZooKeeper, which causes a start failure of ClickHouse.

Solution

Step 1 Back up all data tables in the database of the faulty node to another directory.

- Back up table data:

  cd /srv/BigData/data1/clickhouse/data/*Database name*

  mv *Table name Directory to be backed up/***data***1*

  NOTE

  If there are multiple disks, back up data of **data1** to **dataN**.

- Back up metadata information:

  cd /srv/BigData/data1/clickhouse_path/metadata

  mv *Table name***.sql** *Directory to be backed up*

For example, to back up the lineorder table in the default database to the **/home/backup** directory, run the following command.

cd /srv/BigData/data1/clickhouse/data/default

mv lineorder /home/backup/data1

cd /srv/BigData/data1/clickhouse_path/metadata

mv lineorder.sql /home/backup

Step 2 Log in to MRS Manager, choose **Cluster** > **Services** > **ClickHouse** > **Instance**, select the target instance node, and click **Start Instance**.

Step 3 After the instance is started, use the ClickHouse client to log in to the faulty node.

clickhouse client --host *Clickhouse instance IP address* **--user** *User name* **--password** *Password*

Step 4 Run the following command to obtain the ZooKeeper path **zookeeper_path** of the current table and **replica_num** of the corresponding node.

SELECT zookeeper_path FROM system.replicas WHERE database = '*Database name***' AND table = '***Table name***';**

SELECT replica_num,host_name FROM system.clusters;

Step 5 Run the following command to access the ZooKeeper command line interface:

zkCli.sh -server *IP address of the ZooKeeper node***:2181**

Step 6 Locate the ZooKeeper path corresponding to the table data of the faulty node.

ls *zookeeper_path***/replicas/***replica_num*

NOTE

zookeeper_path indicates the value of **zookeeper_path** obtained in **Step 4**.

replica_num indicates the value of **replica_num** corresponding to the host in **Step 4**.

Step 7 Run the following command to delete the replica data from ZooKeeper:

deleteall *zookeeper_path***/replicas/***replica_num*

Step 8 Use the ClickHouse client to log in to the node and create the ReplicatedMergeTree engine table of the cluster.

clickhouse client --host *Clickhouse instance IP address* **--multiline --user** *Username* **--password** *Password*

CREATE TABLE *Database name.Table name* **ON CLUSTER** *Cluster name*

...

ENGINE = ReplicatedMergeTree ...

The following error message is displayed on other replica nodes, which is normal and can be ignored.

```
Received exception from server (version 20.8.7):
Code: 57. DB::Exception: Received from x.x.x.x:9000. DB::Exception:
There was an error on [x.x.x.x:9000]: Code: 57, e.displayText() =
DB::Exception: Table DEFAULT.lineorder already exists. (version 20.8.11.17
(official build)).
```

After the table is successfully created, the table data on the faulty node will be automatically synchronized. The data restoration is complete.

----End

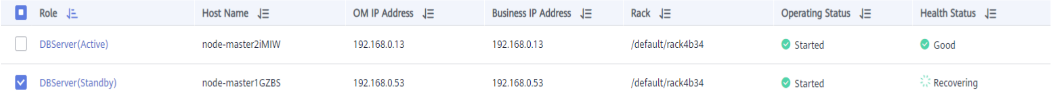
15.5 Using DBService

15.5.1 DBServer Instance Is in Abnormal Status

Symptom

A DBServer instance is in the **Concerning** state for a long period of time.

Figure 15-12 DBServer instance status

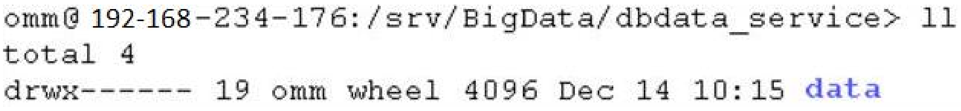


☐ Role	Host Name	OM IP Address	Business IP Address	Rack	Operating Status	Health Status
☐ DBServer(Active)	node-master2iMIW	192.168.0.13	192.168.0.13	/default/rack4b34	✓ Started	✓ Good
☑ DBServer(Standby)	node-master1GZBS	192.168.0.53	192.168.0.53	/default/rack4b34	✓ Started	Recovering

Cause Analysis

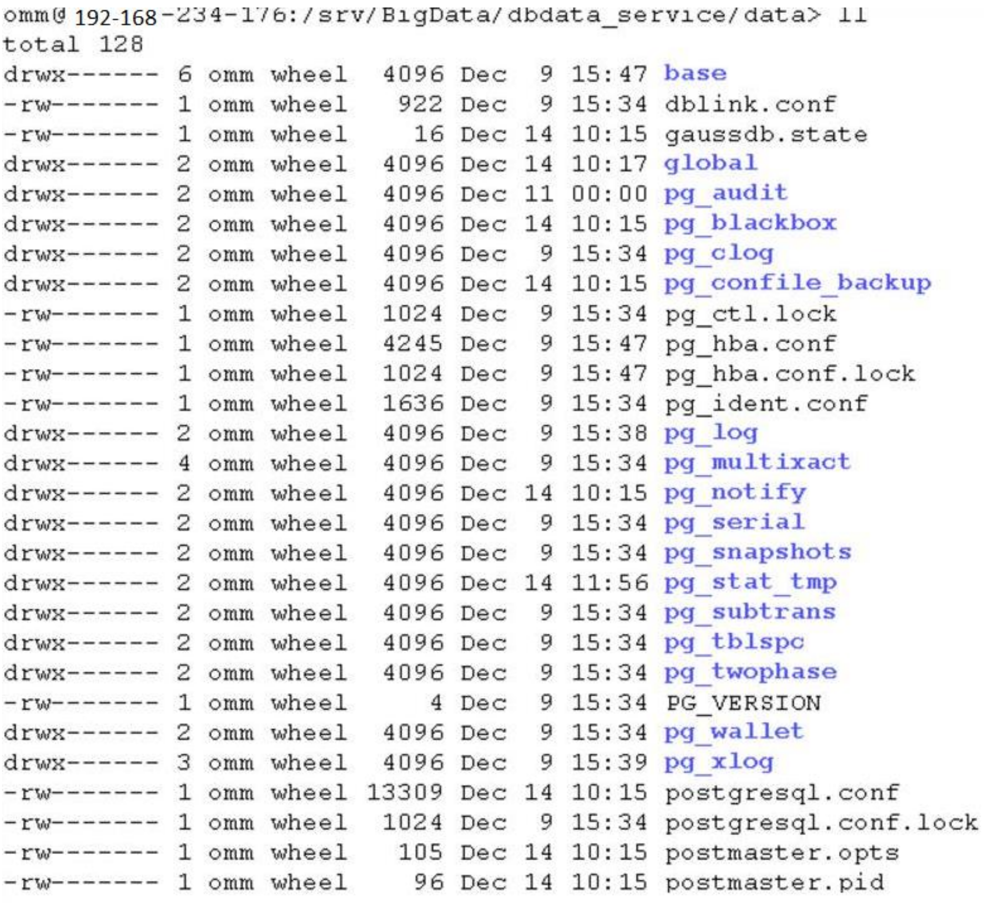
The permission for files or directories in the data directory is incorrect. GaussDB requires that the file permission be at least 600 and directory permission be at least 700.

Figure 15-13 Directory permission list



```
omm@ 192-168-234-176:/srv/BigData/dbdata_service> ll
total 4
drwx------ 19 omm wheel 4096 Dec 14 10:15 data
```

Figure 15-14 File permission list



```
omm@ 192-168-234-176:/srv/BigData/dbdata_service/data> ll
total 128
drwx------ 6 omm wheel  4096 Dec  9 15:47 base
-rw------- 1 omm wheel   922 Dec  9 15:34 dblink.conf
-rw------- 1 omm wheel    16 Dec 14 10:15 gaussdb.state
drwx------ 2 omm wheel  4096 Dec 14 10:17 global
drwx------ 2 omm wheel  4096 Dec 11 00:00 pg_audit
drwx------ 2 omm wheel  4096 Dec 14 10:15 pg_blackbox
drwx------ 2 omm wheel  4096 Dec  9 15:34 pg_clog
drwx------ 2 omm wheel  4096 Dec 14 10:15 pg_confile_backup
-rw------- 1 omm wheel  1024 Dec  9 15:34 pg_ctl.lock
-rw------- 1 omm wheel  4245 Dec  9 15:47 pg_hba.conf
-rw------- 1 omm wheel  1024 Dec  9 15:47 pg_hba.conf.lock
-rw------- 1 omm wheel  1636 Dec  9 15:34 pg_ident.conf
drwx------ 2 omm wheel  4096 Dec  9 15:38 pg_log
drwx------ 4 omm wheel  4096 Dec  9 15:34 pg_multixact
drwx------ 2 omm wheel  4096 Dec 14 10:15 pg_notify
drwx------ 2 omm wheel  4096 Dec  9 15:34 pg_serial
drwx------ 2 omm wheel  4096 Dec  9 15:34 pg_snapshots
drwx------ 2 omm wheel  4096 Dec 14 11:56 pg_stat_tmp
drwx------ 2 omm wheel  4096 Dec  9 15:34 pg_subtrans
drwx------ 2 omm wheel  4096 Dec  9 15:34 pg_tblspc
drwx------ 2 omm wheel  4096 Dec  9 15:34 pg_twophase
-rw------- 1 omm wheel     4 Dec  9 15:34 PG_VERSION
drwx------ 2 omm wheel  4096 Dec  9 15:34 pg_wallet
drwx------ 3 omm wheel  4096 Dec  9 15:39 pg_xlog
-rw------- 1 omm wheel 13309 Dec 14 10:15 postgresql.conf
-rw------- 1 omm wheel  1024 Dec  9 15:34 postgresql.conf.lock
-rw------- 1 omm wheel   105 Dec 14 10:15 postmaster.opts
-rw------- 1 omm wheel    96 Dec 14 10:15 postmaster.pid
```

Solution

Step 1 Modify the permissions on the files and directories based on the permission list in **Figure 15-13** and **Figure 15-14**.

Step 2 Restart the DBServer instance.

----End

15.5.2 DBServer Instance Remains in the Restoring State

Symptom

A DBServer instance remains in the **Restoring** state. The status cannot be recovered even after a restart.

Cause Analysis

1. DBService monitors the **${BIGDATA_HOME}/MRS_***XXX***/install/dbservice/ha/module/harm/plugin/script/gsDB/.startGS.fail** file. *XXX* indicates the product version.
2. If the value in the file is greater than 3, the startup fails. The NodeAgent keeps trying to restart the instance. In this case, the startup still fails and the value is incremented by 1 each time the startup fails.

Solution

Step 1 Log in to MRS Manager.

Step 2 Stop the DBServer instance.

Step 3 Log in to the node where the DBServer instance is abnormal as user **omm**.

Step 4 Change the value of in the **${BIGDATA_HOME}/MRS_***XXX***/install/dbservice/ha/module/harm/plugin/script/gsDB/.startGS.fail** file to **0**. *XXX* indicates the product version.

Step 5 Start the DBServer instance.

----End

15.5.3 Default Port 20050 or 20051 Is Occupied

Symptom

DBService restart fails, and information indicating that port 20050 or 20051 is occupied is displayed in the printed fault log.

Cause Analysis

1. The default port 20050 or 20051 used by DBService is occupied by another process.
2. The DBService process is not stopped, and the port used by DBService is not released.

Solution

This solution uses port 20051 as an example. The solution to the problem that port 20050 is occupied is similar.

Step 1 Log in to the node where the error is reported as user **root**, and run the **netstat -nap | grep 20051** command to check the process that occupies port 20051.

Step 2 Run the **kill** command to forcibly stop the process that uses port 20051.

Step 3 About 2 minutes later, run the **netstat -nap | grep 20051** command again to check whether any process uses the port.

Step 4 Check the service to which the process belongs and change the port for the service.

Step 5 Run the **find . -name "*20051*"** command in the **/tmp** and **/var/run/MRS-DBService/** directories, and delete all files found.

Step 6 Log in to Manager and restart DBService.

----End

15.5.4 DBServer Instance Is Always in the Restoring State Because the Incorrect /tmp Directory Permission

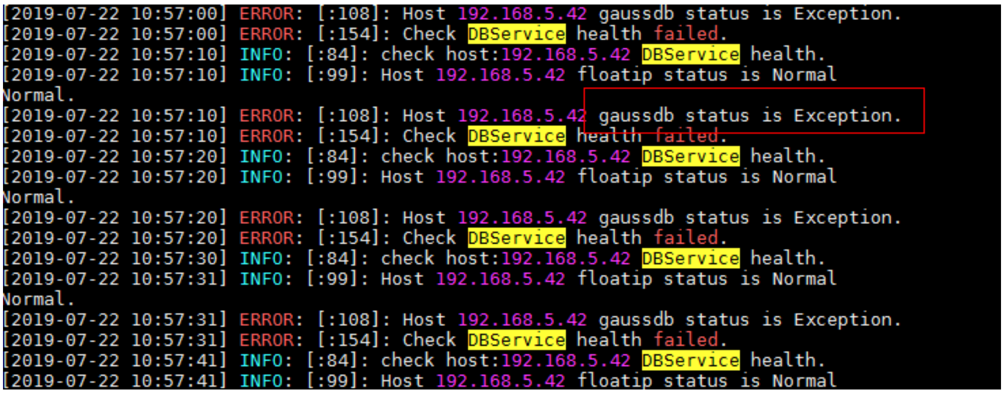
Symptom

A DBServer instance remains in the **Restoring** state. The status cannot be recovered even after a restart.

Cause Analysis

1. Check **/var/log/Bigdata/dbservice/healthCheck/dbservice_processCheck.log**. It is found that GaussDB is abnormal.

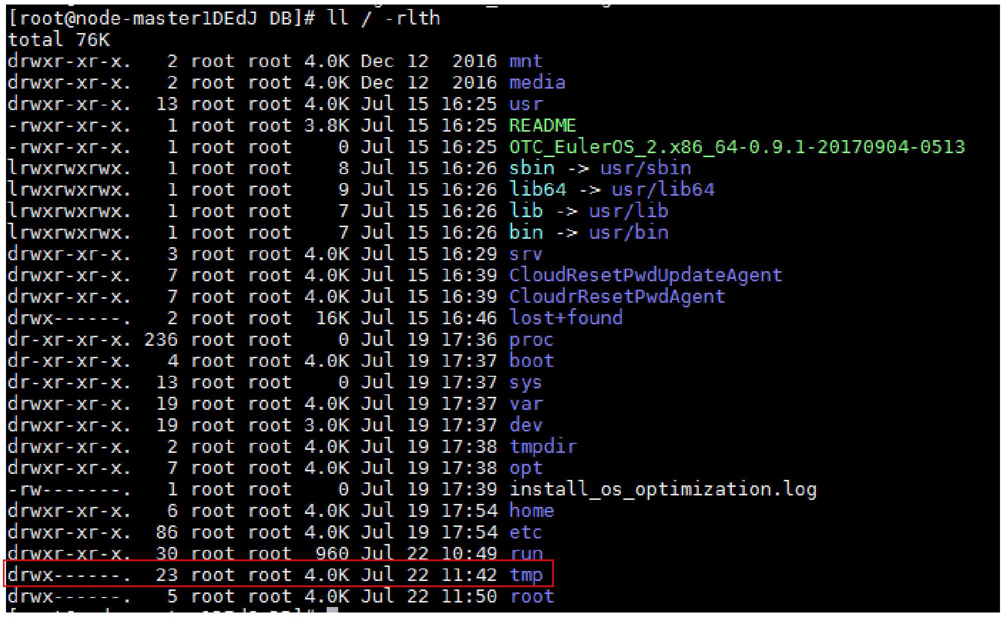
Figure 15-15 GaussDB exception



```
[2019-07-22 10:57:00] ERROR: [:108]: Host 192.168.5.42 gaussdb status is Exception.
[2019-07-22 10:57:00] ERROR: [:154]: Check DBService health failed.
[2019-07-22 10:57:10] INFO: [:84]: check host:192.168.5.42 DBService health.
[2019-07-22 10:57:10] INFO: [:99]: Host 192.168.5.42 floatip status is Normal
Normal.
[2019-07-22 10:57:10] ERROR: [:108]: Host 192.168.5.42 gaussdb status is Exception.
[2019-07-22 10:57:10] ERROR: [:154]: Check DBService health failed.
[2019-07-22 10:57:20] INFO: [:84]: check host:192.168.5.42 DBService health.
[2019-07-22 10:57:20] INFO: [:99]: Host 192.168.5.42 floatip status is Normal
Normal.
[2019-07-22 10:57:20] ERROR: [:108]: Host 192.168.5.42 gaussdb status is Exception.
[2019-07-22 10:57:20] ERROR: [:154]: Check DBService health failed.
[2019-07-22 10:57:30] INFO: [:84]: check host:192.168.5.42 DBService health.
[2019-07-22 10:57:31] INFO: [:99]: Host 192.168.5.42 floatip status is Normal
Normal.
[2019-07-22 10:57:31] ERROR: [:108]: Host 192.168.5.42 gaussdb status is Exception.
[2019-07-22 10:57:31] ERROR: [:154]: Check DBService health failed.
[2019-07-22 10:57:41] INFO: [:84]: check host:192.168.5.42 DBService health.
[2019-07-22 10:57:41] INFO: [:99]: Host 192.168.5.42 floatip status is Normal
```

2. The check result shows that the permission on the **/tmp** directory is incorrect.

Figure 15-16 /tmp permission



```
[root@node-master1DEdJ DB]# ll / -rlth
total 76K
drwxr-xr-x.   2 root root 4.0K Dec 12  2016 mnt
drwxr-xr-x.   2 root root 4.0K Dec 12  2016 media
drwxr-xr-x.  13 root root 4.0K Jul 15 16:25 usr
-rwxr-xr-x.   1 root root 3.8K Jul 15 16:25 README
-rwxr-xr-x.   1 root root    0 Jul 15 16:25 OTC_EulerOS_2.x86_64-0.9.1-20170904-0513
lrwxrwxrwx.   1 root root    8 Jul 15 16:26 sbin -> usr/sbin
lrwxrwxrwx.   1 root root    9 Jul 15 16:26 lib64 -> usr/lib64
lrwxrwxrwx.   1 root root    7 Jul 15 16:26 lib -> usr/lib
lrwxrwxrwx.   1 root root    7 Jul 15 16:26 bin -> usr/bin
drwxr-xr-x.   3 root root 4.0K Jul 15 16:29 srv
drwxr-xr-x.   7 root root 4.0K Jul 15 16:39 CloudResetPwdUpdateAgent
drwxr-xr-x.   7 root root 4.0K Jul 15 16:39 CloudrResetPwdAgent
drwx------.   2 root root  16K Jul 15 16:46 lost+found
dr-xr-xr-x. 236 root root    0 Jul 19 17:36 proc
dr-xr-xr-x.   4 root root 4.0K Jul 19 17:37 boot
dr-xr-xr-x.  13 root root    0 Jul 19 17:37 sys
drwxr-xr-x.  19 root root 4.0K Jul 19 17:37 var
drwxr-xr-x.  19 root root 3.0K Jul 19 17:37 dev
drwxr-xr-x.   2 root root 4.0K Jul 19 17:38 tmpdir
drwxr-xr-x.   7 root root 4.0K Jul 19 17:38 opt
-rw-------.   1 root root    0 Jul 19 17:39 install_os_optimization.log
drwxr-xr-x.   6 root root 4.0K Jul 19 17:54 home
drwxr-xr-x.  86 root root 4.0K Jul 19 17:54 etc
drwxr-xr-x.  30 root root  960 Jul 22 10:49 run
drwx------.  23 root root 4.0K Jul 22 11:42 tmp
drwx------.   5 root root 4.0K Jul 22 11:50 root
```

Solution

Step 1 Run the following command to modify the **/tmp** permission:

chmod 1777 /tmp

Step 2 Wait until the instance status recovers.

----End

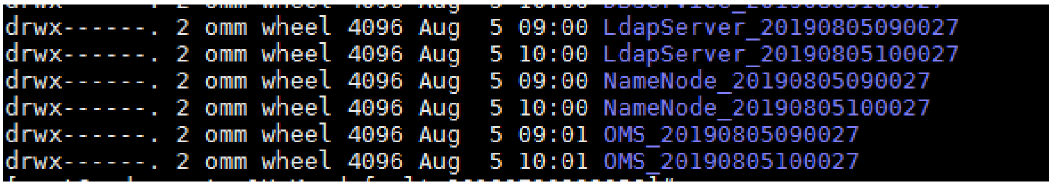
15.5.5 DBService Backup Failure

Symptom

ls /srv/BigData/LocalBackup/default_20190720222358/ -rlth

No DBService backup file exists in the backup file path.

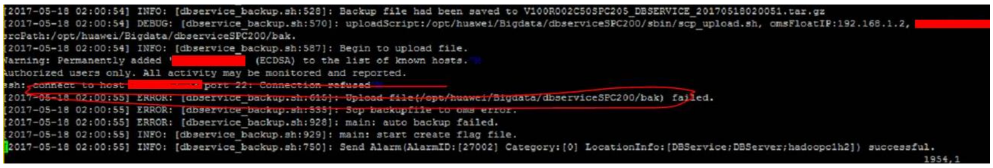
Figure 15-17 Checking the backup file



```
drwx------. 2 omm wheel 4096 Aug  5 09:00 LdapServer_20190805090027
drwx------. 2 omm wheel 4096 Aug  5 10:00 LdapServer_20190805100027
drwx------. 2 omm wheel 4096 Aug  5 09:00 NameNode_20190805090027
drwx------. 2 omm wheel 4096 Aug  5 10:00 NameNode_20190805100027
drwx------. 2 omm wheel 4096 Aug  5 09:01 OMS_20190805090027
drwx------. 2 omm wheel 4096 Aug  5 10:01 OMS_20190805100027
```

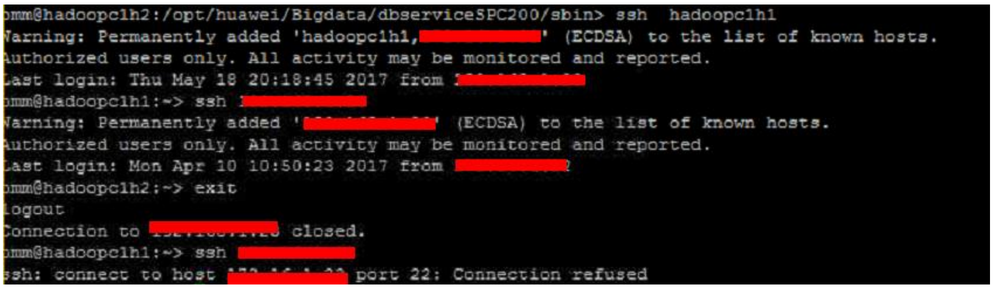
Cause Analysis

- Check the backup log of DBService in **/var/log/Bigdata/dbservice/scriptlog/backup.log**. It is found that the backup is successful but fails to be uploaded to the OMS node.



```
[2017-05-18 02:00:54] INFO: [dbservice_backup.sh:528]: Backup file had been saved to V100R002C50SPC205_DBSERVICE_20170518020051.tar.gz
[2017-05-18 02:00:54] DEBUG: [dbservice_backup.sh:570]: uploadScript:/opt/huawei/Bigdata/dbserviceSPC200/sbin/scp_upload.sh, omsFloatIP:192.168.1.2, [redacted]
srcPath:/opt/huawei/Bigdata/dbserviceSPC200/bak.
[2017-05-18 02:00:54] INFO: [dbservice_backup.sh:587]: Begin to upload file.
Warning: Permanently added '[redacted] (ECDSA) to the list of known hosts.
Authorized users only. All activity may be monitored and reported.
ssh: connect to host [redacted] port 22: Connection refused
[2017-05-18 02:00:55] ERROR: [dbservice_backup.sh:615]: Upload file(/opt/huawei/Bigdata/dbserviceSPC200/bak) failed.
[2017-05-18 02:00:55] ERROR: [dbservice_backup.sh:535]: Scp backupfile to oms error.
[2017-05-18 02:00:55] ERROR: [dbservice_backup.sh:928]: main: auto backup failed.
[2017-05-18 02:00:55] INFO: [dbservice_backup.sh:929]: main: start create flag file.
[2017-05-18 02:00:55] INFO: [dbservice_backup.sh:750]: Send Alarm(AlarmID:[27002] Category:[0] LocationInfo:[DBService;DBServer;hadoopc1h2]) successful.
                                                                                                                            1954,1
```

- The failure is caused by the SSH failure.



```
omm@hadoopc1h2:/opt/huawei/Bigdata/dbserviceSPC200/sbin> ssh  hadoopc1h1
Warning: Permanently added 'hadoopc1h1,[redacted]' (ECDSA) to the list of known hosts.
Authorized users only. All activity may be monitored and reported.
Last login: Thu May 18 20:18:45 2017 from [redacted]
omm@hadoopc1h1:~> ssh [redacted]
Warning: Permanently added '[redacted]' (ECDSA) to the list of known hosts.
Authorized users only. All activity may be monitored and reported.
Last login: Mon Apr 10 10:50:23 2017 from [redacted]
omm@hadoopc1h2:~> exit
logout
Connection to [redacted] closed.
omm@hadoopc1h1:~> ssh [redacted]
ssh: connect to host [redacted] port 22: Connection refused
```

Solution

Step 1 If the network is faulty, contact network engineers.

Step 2 Perform backup operations again after the network fault is rectified.

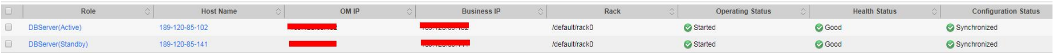
----End

15.5.6 Components Failed to Connect to DBService in Normal State

Symptom

Upper-layer components fail to connect to DBService. The DBService component and two instances are normal.

Figure 15-18 DBService status



□	Role	Host Name	OM IP	Business IP	Rack	Operating Status	Health Status	Configuration Status
□	DBServer(Active)	189-120-85-102	[redacted]	[redacted]	/default/rack0	✓ Started	✓ Good	✓ Synchronized
□	DBServer(Standby)	189-120-85-141	[redacted]	[redacted]	/default/rack0	✓ Started	✓ Good	✓ Synchronized

Cause Analysis

1. The upper-layer component is DBService connected through **dbservice.floatip**.
2. Run the **netstat -anp | grep 20051** command on the node where DBServer resides. It is found that the Gauss process of DBService is not bound to the floating IP address during startup, and only the local IP address 127.0.0.1 is listened.

Solution

Step 1 Restart the DBService service.

Step 2 Run the **netstat -anp | grep 20051** command on the active DBServer node to check whether **dbservice.floatip** is bound.

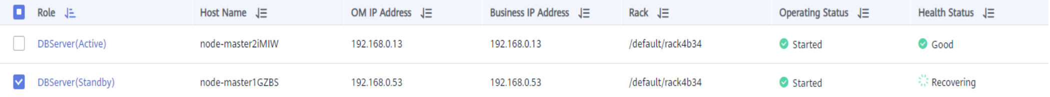
----End

15.5.7 DBServer Failed to Start

Symptom

DBService fails to be started and restarts also fail. The instance keeps in the **Recovering** state.

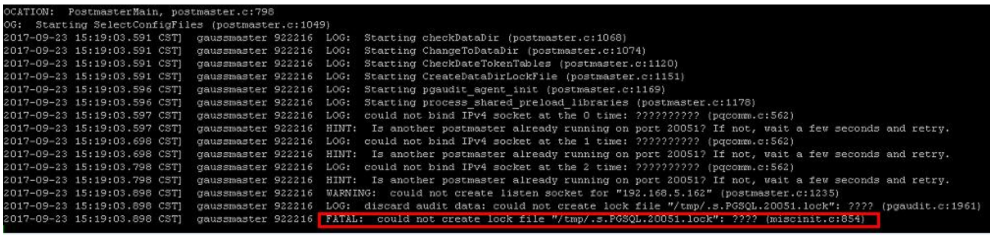
Figure 15-19 DBService status



■ Role	Host Name	OM IP Address	Business IP Address	Rack	Operating Status	Health Status
□ DBServer(Active)	node-master2iMIW	192.168.0.13	192.168.0.13	/default/rack4b34	✓ Started	✓ Good
☑ DBServer(Standby)	node-master1GZBS	192.168.0.53	192.168.0.53	/default/rack4b34	✓ Started	Recovering

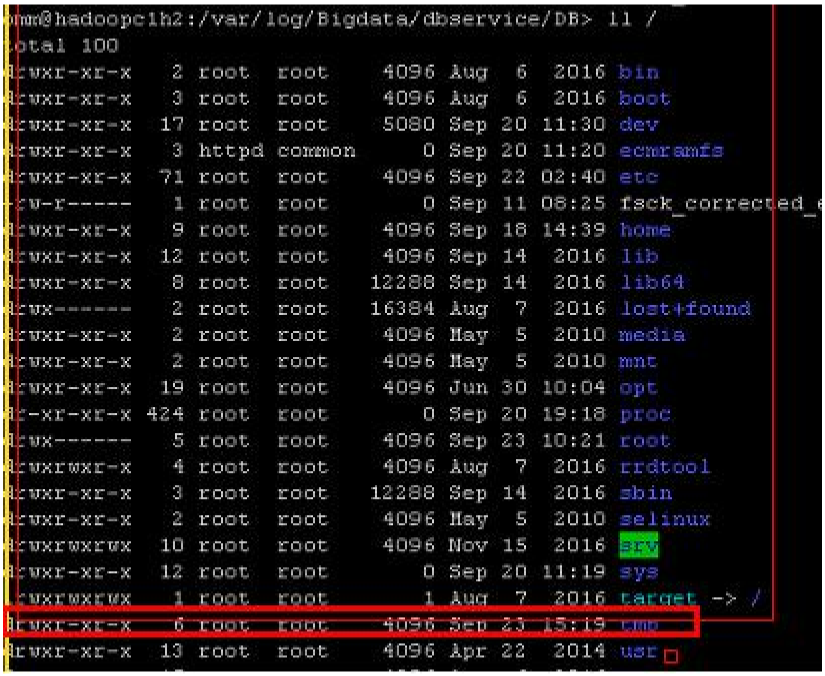
Cause Analysis

1. Check the DBService logs in **/var/log/Bigdata/dbservice/DB/gs_ctl-current.log**. The following error message is displayed:



```
OCATION:  PostmasterMain, postmaster.c:798
OG:  Starting SelectConfigFiles (postmaster.c:1049)
2017-09-23 15:19:03.591 CST]  gaussmaster 922216  LOG:  Starting checkDataDir (postmaster.c:1068)
2017-09-23 15:19:03.591 CST]  gaussmaster 922216  LOG:  Starting ChangeToDataDir (postmaster.c:1074)
2017-09-23 15:19:03.591 CST]  gaussmaster 922216  LOG:  Starting CheckDateTokenTables (postmaster.c:1120)
2017-09-23 15:19:03.591 CST]  gaussmaster 922216  LOG:  Starting CreateDataDirLockFile (postmaster.c:1151)
2017-09-23 15:19:03.596 CST]  gaussmaster 922216  LOG:  Starting pgaudit_agent_init (postmaster.c:1169)
2017-09-23 15:19:03.596 CST]  gaussmaster 922216  LOG:  Starting process_shared_preload_libraries (postmaster.c:1178)
2017-09-23 15:19:03.597 CST]  gaussmaster 922216  LOG:  could not bind IPv4 socket at the 0 time: ?????????? (pqcomm.c:562)
2017-09-23 15:19:03.597 CST]  gaussmaster 922216  HINT:  Is another postmaster already running on port 20051? If not, wait a few seconds and retry.
2017-09-23 15:19:03.698 CST]  gaussmaster 922216  LOG:  could not bind IPv4 socket at the 1 time: ?????????? (pqcomm.c:562)
2017-09-23 15:19:03.698 CST]  gaussmaster 922216  HINT:  Is another postmaster already running on port 20051? If not, wait a few seconds and retry.
2017-09-23 15:19:03.798 CST]  gaussmaster 922216  LOG:  could not bind IPv4 socket at the 2 time: ?????????? (pqcomm.c:562)
2017-09-23 15:19:03.798 CST]  gaussmaster 922216  HINT:  Is another postmaster already running on port 20051? If not, wait a few seconds and retry.
2017-09-23 15:19:03.898 CST]  gaussmaster 922216  WARNING:  could not create listen socket for "192.168.5.162" (postmaster.c:1235)
2017-09-23 15:19:03.898 CST]  gaussmaster 922216  LOG:  discard audit data: could not create lock file "/tmp/.s.PGSQL.20051.lock": ???? (pgaudit.c:1961)
2017-09-23 15:19:03.898 CST]  gaussmaster 922216  FATAL:  could not create lock file "/tmp/.s.PGSQL.20051.lock": ???? (miscinit.c:854)
```

2. It is found that the **/tmp** permission is incorrect. The correct value should be **777**.



```
omm@hadoopc1h2:/var/log/Bigdata/dbservice/DB> ll /
total 100
drwxr-xr-x   2 root  root     4096 Aug  6  2016 bin
drwxr-xr-x   3 root  root     4096 Aug  6  2016 boot
drwxr-xr-x  17 root  root     5080 Sep 20 11:30 dev
drwxr-xr-x   3 httpd common      0 Sep 20 11:20 ecmramfs
drwxr-xr-x  71 root  root     4096 Sep 22 02:40 etc
-rw-r-----   1 root  root        0 Sep 11 08:25 fsck_corrected_e
drwxr-xr-x   9 root  root     4096 Sep 18 14:39 home
drwxr-xr-x  12 root  root     4096 Sep 14  2016 lib
drwxr-xr-x   8 root  root    12288 Sep 14  2016 lib64
drwx------   2 root  root    16384 Aug  7  2016 lost+found
drwxr-xr-x   2 root  root     4096 May  5  2010 media
drwxr-xr-x   2 root  root     4096 May  5  2010 mnt
drwxr-xr-x  19 root  root     4096 Jun 30 10:04 opt
dr-xr-xr-x 424 root  root        0 Sep 20 19:18 proc
drwx------   5 root  root     4096 Sep 23 10:21 root
drwxrwxr-x   4 root  root     4096 Aug  7  2016 rrdtool
drwxr-xr-x   3 root  root    12288 Sep 14  2016 sbin
drwxr-xr-x   2 root  root     4096 May  5  2010 selinux
drwxrwxrwx  10 root  root     4096 Nov 15  2016 srv
drwxr-xr-x  12 root  root        0 Sep 20 11:19 sys
lrwxrwxrwx   1 root  root        1 Aug  7  2016 target -> /
drwxr-xr-x   6 root  root     4096 Sep 23 15:19 tmp
drwxr-xr-x  13 root  root     4096 Apr 22  2014 usr
```

Solution

Step 1 Modify the **/tmp** permission by changing the value to **777**.

Step 2 Restart DBService.

----End

15.5.8 DBService Backup Failed Because the Floating IP Address Is Unreachable

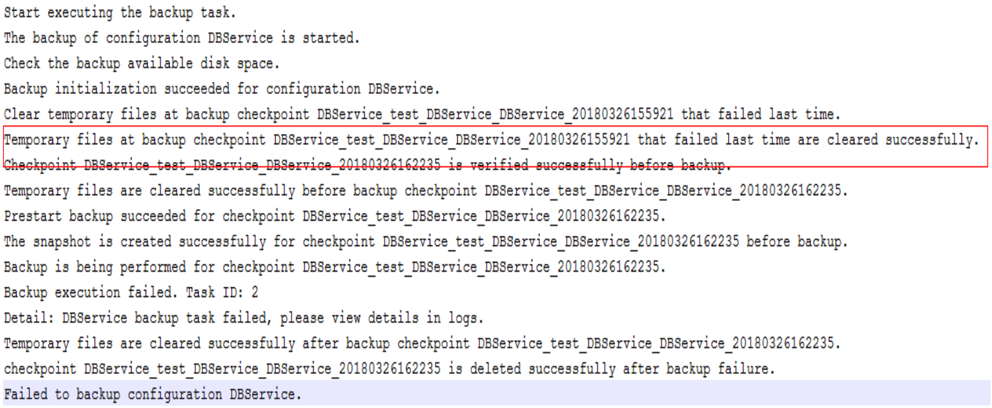
Symptom

The default DBService backup fails, but backups of NameNode, LdapServer, and OMS are successful.

Cause Analysis

1. Check the error information on the DBService backup page:

   ```
   Clear temporary files at backup checkpoint DBService_test_DBService_DBService_20180326155921
   that fialed last time.
   Temporoyary files at backup checkpoint DBService_test_DBService_DBService20180326155921 that
   failed last time are cleared successfully.
   ```

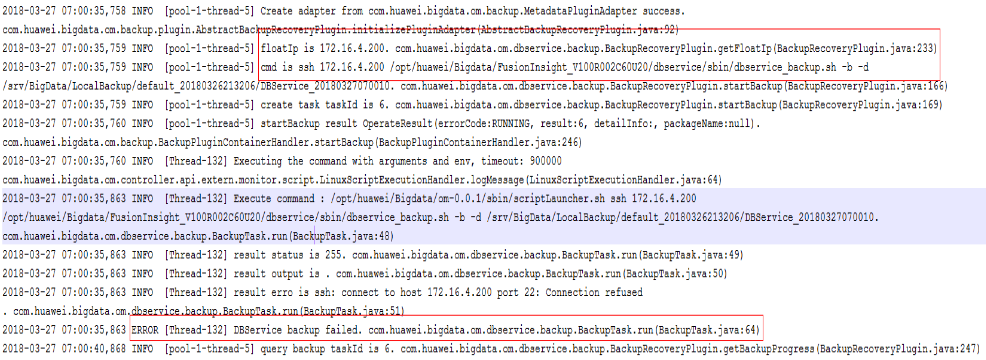
   

   ```
   Start executing the backup task.
   The backup of configuration DBService is started.
   Check the backup available disk space.
   Backup initialization succeeded for configuration DBService.
   Clear temporary files at backup checkpoint DBService_test_DBService_DBService_20180326155921 that failed last time.
   Temporary files at backup checkpoint DBService_test_DBService_DBService_20180326155921 that failed last time are cleared successfully.
   Checkpoint DBService_test_DBService_DBService_20180326162235 is verified successfully before backup.
   Temporary files are cleared successfully before backup checkpoint DBService_test_DBService_DBService_20180326162235.
   Prestart backup succeeded for checkpoint DBService_test_DBService_DBService_20180326162235.
   The snapshot is created successfully for checkpoint DBService_test_DBService_DBService_20180326162235 before backup.
   Backup is being performed for checkpoint DBService_test_DBService_DBService_20180326162235.
   Backup execution failed. Task ID: 2
   Detail: DBService backup task failed, please view details in logs.
   Temporary files are cleared successfully after backup checkpoint DBService_test_DBService_DBService_20180326162235.
   checkpoint DBService_test_DBService_DBService_20180326162235 is deleted successfully after backup failure.
   Failed to backup configuration DBService.
   ```

2. Check the **/var/log/Bigdata/dbservice/scriptlog/backup.log** file. It is found that the log printing stops and no related backup information is found.

3. Check the **/var/log/Bigdata/controller/backupplugin.log** file on the active OMS node. The following error information is found:

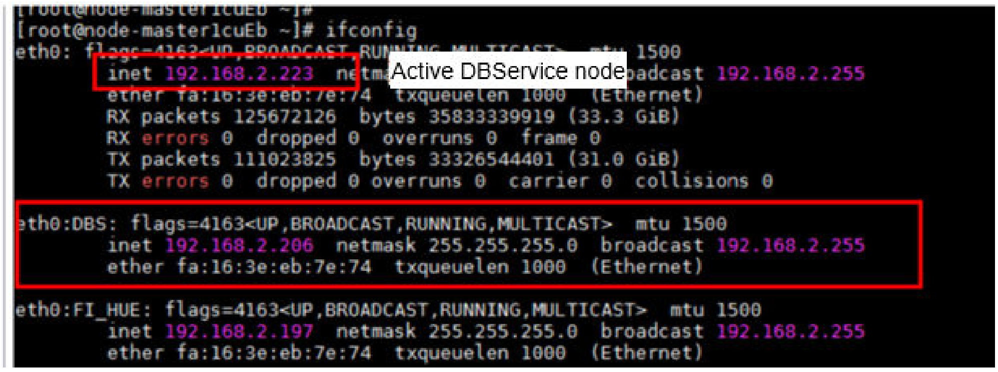
   ```
   result error is ssh:connect to host 172.16.4.200 port 22: Connection refused (172.16.4.200 is the
   floating IP address of DBService)
   DBService backup failed.
   ```

   ```
   2018-03-27 07:00:35,758 INFO  [pool-1-thread-5] Create adapter from com.huawei.bigdata.om.backup.MetadataPluginAdapter success.
   com.huawei.bigdata.om.backup.plugin.AbstractBackupRecoveryPlugin.initializePluginAdapter(AbstractBackupRecoveryPlugin.java:92)
   2018-03-27 07:00:35,759 INFO  [pool-1-thread-5] floatIp is 172.16.4.200. com.huawei.bigdata.om.dbservice.backup.BackupRecoveryPlugin.getFloatIp(BackupRecoveryPlugin.java:233)
   2018-03-27 07:00:35,759 INFO  [pool-1-thread-5] cmd is ssh 172.16.4.200 /opt/huawei/Bigdata/FusionInsight_V100R002C60U20/dbservice/sbin/dbservice_backup.sh -b -d
   /srv/BigData/LocalBackup/default_20180326213206/DBService_20180327070010. com.huawei.bigdata.om.dbservice.backup.BackupRecoveryPlugin.startBackup(BackupRecoveryPlugin.java:166)
   2018-03-27 07:00:35,759 INFO  [pool-1-thread-5] create task taskId is 6. com.huawei.bigdata.om.dbservice.backup.BackupRecoveryPlugin.startBackup(BackupRecoveryPlugin.java:169)
   2018-03-27 07:00:35,760 INFO  [pool-1-thread-5] startBackup result OperateResult(errorCode:RUNNING, result:6, detailInfo:, packageName:null).
   com.huawei.bigdata.om.backup.BackupPluginContainerHandler.startBackup(BackupPluginContainerHandler.java:246)
   2018-03-27 07:00:35,760 INFO  [Thread-132] Executing the command with arguments and env, timeout: 900000
   com.huawei.bigdata.om.controller.api.extern.monitor.script.LinuxScriptExecutionHandler.logMessage(LinuxScriptExecutionHandler.java:64)
   2018-03-27 07:00:35,863 INFO  [Thread-132] Execute command : /opt/huawei/Bigdata/om-0.0.1/sbin/scriptLauncher.sh ssh 172.16.4.200
   /opt/huawei/Bigdata/FusionInsight_V100R002C60U20/dbservice/sbin/dbservice_backup.sh -b -d /srv/BigData/LocalBackup/default_20180326213206/DBService_20180327070010.
   com.huawei.bigdata.om.dbservice.backup.BackupTask.run(BackupTask.java:48)
   2018-03-27 07:00:35,863 INFO  [Thread-132] result status is 255. com.huawei.bigdata.om.dbservice.backup.BackupTask.run(BackupTask.java:49)
   2018-03-27 07:00:35,863 INFO  [Thread-132] result output is . com.huawei.bigdata.om.dbservice.backup.BackupTask.run(BackupTask.java:50)
   2018-03-27 07:00:35,863 INFO  [Thread-132] result erro is ssh: connect to host 172.16.4.200 port 22: Connection refused
   . com.huawei.bigdata.om.dbservice.backup.BackupTask.run(BackupTask.java:51)
   2018-03-27 07:00:35,863 ERROR [Thread-132] DBService backup failed. com.huawei.bigdata.om.dbservice.backup.BackupTask.run(BackupTask.java:64)
   2018-03-27 07:00:40,868 INFO  [pool-1-thread-5] query backup taskId is 6. com.huawei.bigdata.om.dbservice.backup.BackupRecoveryPlugin.getBackupProgress(BackupRecoveryPlugin.java:247)
   ```

Solution

Step 1 Log in to the active DBService node (the Master node bound with the DBService floating IP address).



```
[root@node-master1cuEb ~]#
[root@node-master1cuEb ~]# ifconfig
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.2.223  netm                     adcast 192.168.2.255
        ether fa:16:3e:eb:7e:74  txqueuelen 1000  (Ethernet)
        RX packets 125672126  bytes 35833339919 (33.3 GiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 111023825  bytes 33326544401 (31.0 GiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

eth0:DBS: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.2.206  netmask 255.255.255.0  broadcast 192.168.2.255
        ether fa:16:3e:eb:7e:74  txqueuelen 1000  (Ethernet)

eth0:FI_HUE: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.2.197  netmask 255.255.255.0  broadcast 192.168.2.255
        ether fa:16:3e:eb:7e:74  txqueuelen 1000  (Ethernet)
```

Step 2 Add the DBService floating IP address to **ListenAddress** or comment out **ListenAddress** in the **/etc/ssh/sshd_config** file.

Step 3 Run the following command to restart the SSHD service:

service sshd restart

Step 4 Check whether the next DBServie backup is successful.

----End

15.5.9 DBService Failed to Start Due to the Loss of the DBService Configuration File

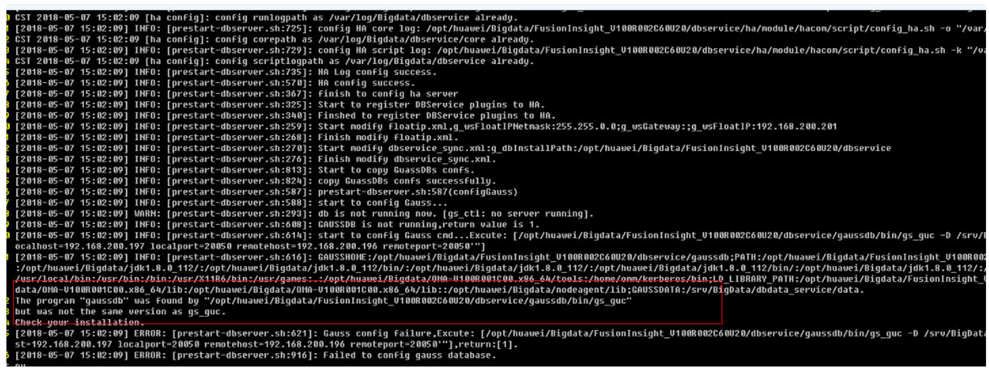
Symptom

The nodes are powered off unexpectedly, and the standby DBService node fails to be restarted.

Cause Analysis

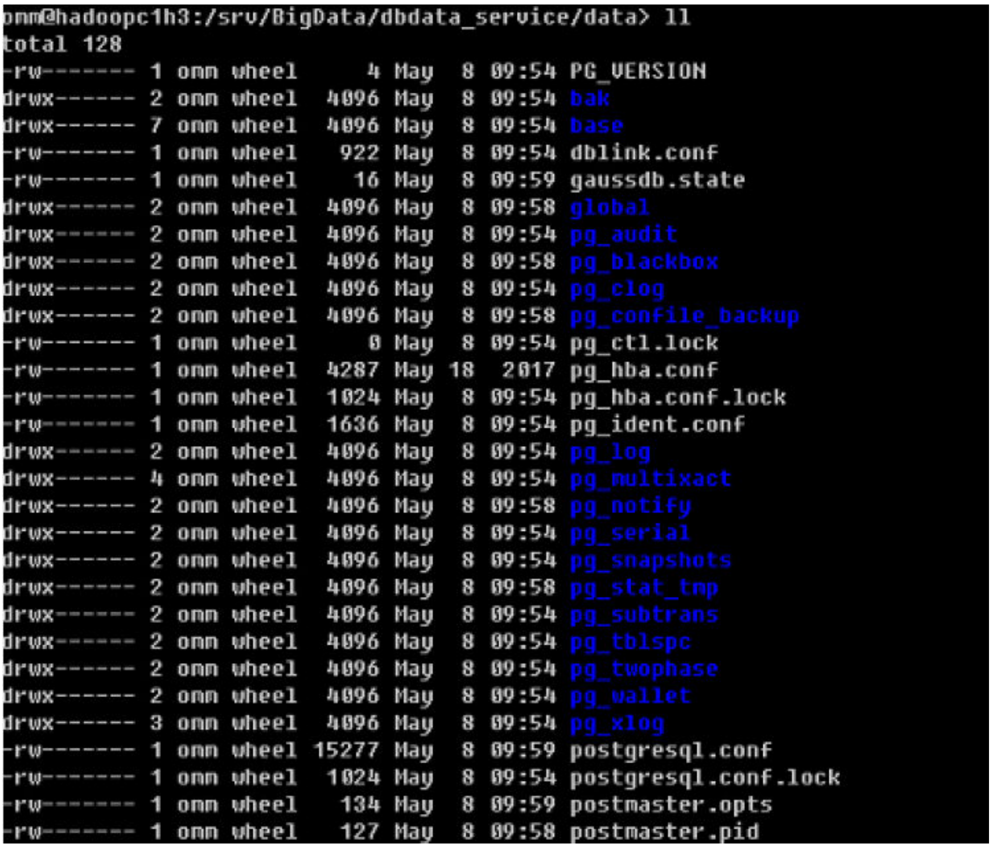
1. The **/var/log/Bigdata/dbservice/DB/gaussdb.log** file is viewed, which contains no information.
2. The **/var/log/Bigdata/dbservice/scriptlog/preStartDBService.log** file is viewed. This file contains the following information, indicating that the configuration information is lost:

```
The program "gaussdb" was found by "
/opt/Bigdata/MRS_xxx/install/dbservice/gaussdb/bin/gs_guc)
But not was not the same version as gs_guc.
Check your installation.
```

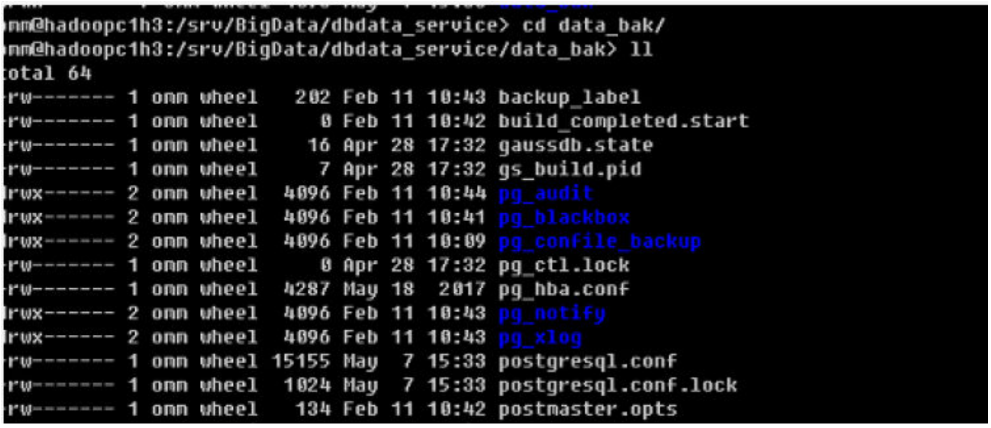


```
CST 2018-05-07 15:02:09 [ha config]: config runlogpath as /var/log/Bigdata/dbservice already.
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:725]: config HA core log: /opt/huawei/Bigdata/FusionInsight_V100R002C60U20/dbservice/ha/module/hacom/script/config_ha.sh -o "/var/
CST 2018-05-07 15:02:09 [ha config]: config corepath as /var/log/Bigdata/dbservice/core already.
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:729]: config HA script log: /opt/huawei/Bigdata/FusionInsight_V100R002C60U20/dbservice/ha/module/hacom/script/config_ha.sh -k "/va
CST 2018-05-07 15:02:09 [ha config]: config scriptlogpath as /var/log/Bigdata/dbservice already.
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:735]: HA Log config success.
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:570]: HA config success.
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:367]: finish to config ha server
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:325]: Start to register DBService plugins to HA.
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:340]: Finshed to register DBService plugins to HA.
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:259]: Start modify floatip.xml,g_wsFloatIPNetmask:255.255.0.0;g_wsGateway:;g_wsFloatIP:192.168.200.201
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:268]: Finish modify floatip.xml.
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:270]: Start modify dbservice_sync.xml:g_dbInstallPath:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/dbservice
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:276]: Finish modify dbservice_sync.xml.
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:813]: Start to copy GuassDBs confs.
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:824]: copy GuassDBs confs successfully.
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:587]: prestart-dbserver.sh:587(configGauss)
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:588]: start to config Gauss...
[2018-05-07 15:02:09] WARN: [prestart-dbserver.sh:293]: db is not running now. [gs_ctl: no server running].
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:608]: GAUSSDB is not running,return value is 1.
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:614]: start to config Gauss cmd...Excute: [/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/dbservice/gaussdb/bin/gs_guc -D /srv/B
ocalhost=192.168.200.197 localport=20050 remotehost=192.168.200.196 remoteport=20050'"]
[2018-05-07 15:02:09] INFO: [prestart-dbserver.sh:616]: GAUSSHOME:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/dbservice/gaussdb;PATH:/opt/huawei/Bigdata/FusionInsight_V100R002
:/opt/huawei/Bigdata/jdk1.8.0_112/:/opt/huawei/Bigdata/jdk1.8.0_112/bin/:/opt/huawei/Bigdata/jdk1.8.0_112/:/opt/huawei/Bigdata/jdk1.8.0_112/bin/:/opt/huawei/Bigdata/jdk1.8.0_112/:/
/usr/local/bin:/usr/bin:/bin:/usr/X11R6/bin:/usr/games:.:/opt/huawei/Bigdata/OMA-V100R001C00.x86_64/tools:/home/omm/kerberos/bin;LD_LIBRARY_PATH:/opt/huawei/Bigdata/FusionInsight_V
data/OMA-V100R001C00.x86_64/lib:/opt/huawei/Bigdata/OMA-V100R001C00.x86_64/lib::/opt/huawei/Bigdata/nodeagent/lib;GAUSSDATA:/srv/BigData/dbdata_service/data.
The program "gaussdb" was found by "/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/dbservice/gaussdb/bin/gs_guc"
but was not the same version as gs_guc.
Check your installation.
[2018-05-07 15:02:09] ERROR: [prestart-dbserver.sh:621]: Gauss config failure,Excute: [/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/dbservice/gaussdb/bin/gs_guc -D /srv/BigData
st=192.168.200.197 localport=20050 remotehost=192.168.200.196 remoteport=20050'"],return:[1].
[2018-05-07 15:02:09] ERROR: [prestart-dbserver.sh:916]: Failed to config gauss database.
```

3. The configuration file in the **/srv/BigData/dbdata_service/data** directory on the active DBServer node is compared with the configuration file in the **/srv/BigData/dbdata_service/data** directory on the standby DBServer node, which shows major difference.



```
omm@hadoopc1h3:/srv/BigData/dbdata_service/data> ll
total 128
-rw------- 1 omm wheel     4 May  8 09:54 PG_VERSION
drwx------ 2 omm wheel  4096 May  8 09:54 bak
drwx------ 7 omm wheel  4096 May  8 09:54 base
-rw------- 1 omm wheel   922 May  8 09:54 dblink.conf
-rw------- 1 omm wheel    16 May  8 09:59 gaussdb.state
drwx------ 2 omm wheel  4096 May  8 09:58 global
drwx------ 2 omm wheel  4096 May  8 09:54 pg_audit
drwx------ 2 omm wheel  4096 May  8 09:58 pg_blackbox
drwx------ 2 omm wheel  4096 May  8 09:54 pg_clog
drwx------ 2 omm wheel  4096 May  8 09:58 pg_confile_backup
-rw------- 1 omm wheel     0 May  8 09:54 pg_ctl.lock
-rw------- 1 omm wheel  4287 May 18  2017 pg_hba.conf
-rw------- 1 omm wheel  1024 May  8 09:54 pg_hba.conf.lock
-rw------- 1 omm wheel  1636 May  8 09:54 pg_ident.conf
drwx------ 2 omm wheel  4096 May  8 09:54 pg_log
drwx------ 4 omm wheel  4096 May  8 09:54 pg_multixact
drwx------ 2 omm wheel  4096 May  8 09:58 pg_notify
drwx------ 2 omm wheel  4096 May  8 09:54 pg_serial
drwx------ 2 omm wheel  4096 May  8 09:54 pg_snapshots
drwx------ 2 omm wheel  4096 May  8 09:58 pg_stat_tmp
drwx------ 2 omm wheel  4096 May  8 09:54 pg_subtrans
drwx------ 2 omm wheel  4096 May  8 09:54 pg_tblspc
drwx------ 2 omm wheel  4096 May  8 09:54 pg_twophase
drwx------ 2 omm wheel  4096 May  8 09:54 pg_wallet
drwx------ 3 omm wheel  4096 May  8 09:54 pg_xlog
-rw------- 1 omm wheel 15277 May  8 09:59 postgresql.conf
-rw------- 1 omm wheel  1024 May  8 09:54 postgresql.conf.lock
-rw------- 1 omm wheel   134 May  8 09:59 postmaster.opts
-rw------- 1 omm wheel   127 May  8 09:58 postmaster.pid
```

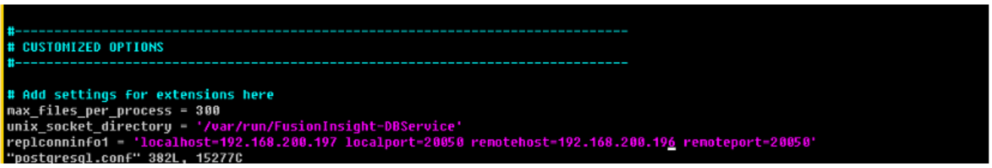


```
mm@hadoopc1h3:/srv/BigData/dbdata_service> cd data_bak/
mm@hadoopc1h3:/srv/BigData/dbdata_service/data_bak> ll
otal 64
rw------- 1 omm wheel   202 Feb 11 10:43 backup_label
rw------- 1 omm wheel     0 Feb 11 10:42 build_completed.start
rw------- 1 omm wheel    16 Apr 28 17:32 gaussdb.state
rw------- 1 omm wheel     7 Apr 28 17:32 gs_build.pid
rwx------ 2 omm wheel  4096 Feb 11 10:44 pg_audit
rwx------ 2 omm wheel  4096 Feb 11 10:41 pg_blackbox
rwx------ 2 omm wheel  4096 Feb 11 10:09 pg_confile_backup
rw------- 1 omm wheel     0 Apr 28 17:32 pg_ctl.lock
rw------- 1 omm wheel  4287 May 18  2017 pg_hba.conf
rwx------ 2 omm wheel  4096 Feb 11 10:43 pg_notify
rwx------ 2 omm wheel  4096 Feb 11 10:43 pg_xlog
rw------- 1 omm wheel 15155 May  7 15:33 postgresql.conf
rw------- 1 omm wheel  1024 May  7 15:33 postgresql.conf.lock
rw------- 1 omm wheel   134 Feb 11 10:42 postmaster.opts
```

Solution

Step 1 Copy the content in the **/srv/BigData/dbdata_service/data** directory on the active node to the standby node and ensure that the file permission and owner group are the same as those on the active node.

Step 2 Modify configuration in **postgresql.conf**. Set **localhost** to the IP of the local node and **remotehost** to the IP of the peer node.



```
#------------------------------------------------------------------------------
# CUSTOMIZED OPTIONS
#------------------------------------------------------------------------------

# Add settings for extensions here
max_files_per_process = 300
unix_socket_directory = '/var/run/FusionInsight-DBService'
replconninfo1 = 'localhost=192.168.200.197 localport=20050 remotehost=192.168.200.196 remoteport=20050'
"postgresql.conf" 382L, 15277C
```

Step 3 Log in to Manager and restart the standby DBServer node.

----End

15.6 Using Flink

15.6.1 "IllegalConfigurationException: Error while parsing YAML configuration file: "security.kerberos.login.keytab" Is Displayed When a Command Is Executed on an Installed Client

Symptom

After the client is successfully installed, an error message "IllegalConfigurationException: Error while parsing YAML configuration file:"security.kerberos.login.keytab" is displayed when the command (for example, **yarn-session.sh**) on the client is executed.

```
[root@8-5-131-10 bin]# yarn-session.sh
2018-10-25 01:22:06,454 | ERROR | [main] | Error while trying to split key and value in configuration
file /opt/flinkclient/Flink/flink/conf/flink-conf.yaml:80: "security.kerberos.login.keytab: " |
org.apache.flink.configuration.GlobalConfiguration (GlobalConfiguration.java:160)
Exception in thread "main" org.apache.flink.configuration.IllegalConfigurationException: Error while parsing
YAML configuration file :80: "security.kerberos.login.keytab: "
```

```
        at org.apache.flink.configuration.GlobalConfiguration.loadYAMLResource(GlobalConfiguration.java:
161)
        at org.apache.flink.configuration.GlobalConfiguration.loadConfiguration(GlobalConfiguration.java:112)
        at org.apache.flink.configuration.GlobalConfiguration.loadConfiguration(GlobalConfiguration.java:79)
        at org.apache.flink.yarn.cli.FlinkYarnSessionCli.main(FlinkYarnSessionCli.java:482)
[root@8-5-131-10 bin]#
```

Cause Analysis

In a secure cluster environment, Flink requires security authentication. The security authentication is not configured on the current client.

1. The following two authentication modes are available for Flink.
   - Kerberos authentication: Flink Yarn client, Yarn ResourceManager, JobManager, HDFS, TaskManager, Kafka, and ZooKeeper
   - Internal authentication mechanism of Yarn: The internal authentication used between YarnResource Manager and Application Master (AM).
2. If a security cluster is required, the Kerberos authentication and security cookie authentication are mandatory. As shown in the logs, it is found that the **security.kerberos.login.keytab** setting in the configuration file is incorrect and the security configuration is not performed.

Solution

Step 1 Download the keytab file from MRS and save it in a folder on a host where the Flink client resides.

Step 2 Configure following parameters in the **flink-conf.yaml** file:

1. Keytab path

   ```
   security.kerberos.login.keytab: /home/flinkuser/keytab/abc222.keytab
   ```

   NOTE

   - **/home/flinkuser/keytab/abc222.keytab** indicates the user directory, which is the directory saves the keytab file in **Step 1**.
   - Ensure that the client user has the permission on the corresponding directory.

2. Principal name

   ```
   security.kerberos.login.principal: abc222
   ```

3. In HA mode, if Zookeeper is configured, the ZooKeeper Kerberos authentication configuration items must be configured as follows:

   ```
   zookeeper.sasl.disable: false
   security.kerberos.login.contexts: Client
   ```

4. If Kerberos authentication is required between the Kafka client and Kafka broker, configure it as follows:

   ```
   security.kerberos.login.contexts: Client,KafkaClient
   ```

----End

15.6.2 "IllegalConfigurationException: Error while parsing YAML configuration file" Is Displayed When a Command Is Executed After Configurations of the Installed Client Are Changed

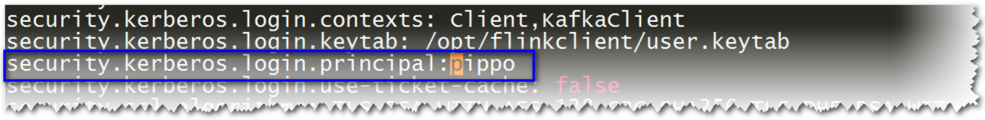
Symptom

After the client is successfully installed, an error message "IllegalConfigurationException: Error while parsing YAML configuration file: 81: "security.kerberos.login.principal:pippo " is displayed when the command (for example, **yarn-session.sh**) on the client is executed.

```
[root@8-5-131-10 bin]# yarn-session.sh
2018-10-25 19:27:01,397 | ERROR | [main] | Error while trying to split key and value in configuration
file /opt/flinkclient/Flink/flink/conf/flink-conf.yaml:81: "security.kerberos.login.principal:pippo " |
org.apache.flink.configuration.GlobalConfiguration  (GlobalConfiguration.java:160)
Exception in thread "main" org.apache.flink.configuration.IllegalConfigurationException: Error while parsing
YAML configuration file :81: "security.kerberos.login.principal:pippo "
        at org.apache.flink.configuration.GlobalConfiguration.loadYAMLResource(GlobalConfiguration.java:
161)
        at org.apache.flink.configuration.GlobalConfiguration.loadConfiguration(GlobalConfiguration.java:112)
        at org.apache.flink.configuration.GlobalConfiguration.loadConfiguration(GlobalConfiguration.java:79)
        at org.apache.flink.yarn.cli.FlinkYarnSessionCli.main(FlinkYarnSessionCli.java:482)
```

Cause Analysis

The **security.kerberos.login.principal:pippo** item in the **flink-conf.yaml** configuration file was faulty.

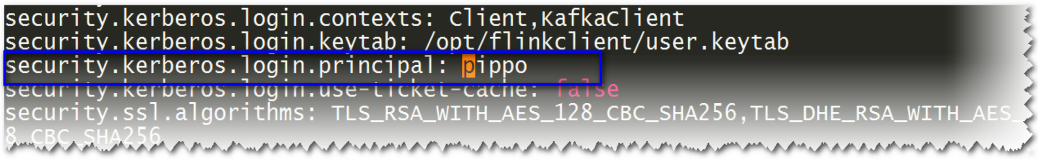


```
security.kerberos.login.contexts: Client,KafkaClient
security.kerberos.login.keytab: /opt/flinkclient/user.keytab
security.kerberos.login.principal:pippo
security.kerberos.login.use-ticket-cache: false
```

Solution

Modify the configuration in the **flink-conf.yaml** file.

Note: The configuration item name and value must be separated by a space.



```
security.kerberos.login.contexts: Client,KafkaClient
security.kerberos.login.keytab: /opt/flinkclient/user.keytab
security.kerberos.login.principal: pippo
security.kerberos.login.use-ticket-cache: false
security.ssl.algorithms: TLS_RSA_WITH_AES_128_CBC_SHA256,TLS_DHE_RSA_WITH_AES_
8_CBC_SHA256
```

15.6.3 The yarn-session.sh Command Fails to Be Executed When the Flink Cluster Is Created

Symptom

During the creation of the Flink cluster, an error message is displayed after the **yarn-session.sh** command execution is suspended.

```
2018-09-20 22:51:16,842 | WARN  | [main] | Unable to get ClusterClient status from Application Client |
org.apache.flink.yarn.YarnClusterClient (YarnClusterClient.java:253)
org.apache.flink.util.FlinkException: Could not connect to the leading JobManager. Please check that the
JobManager is running.
    at org.apache.flink.client.program.ClusterClient.getJobManagerGateway(ClusterClient.java:861)
    at org.apache.flink.yarn.YarnClusterClient.getClusterStatus(YarnClusterClient.java:248)
    at org.apache.flink.yarn.YarnClusterClient.waitForClusterToBeReady(YarnClusterClient.java:516)
    at org.apache.flink.yarn.cli.FlinkYarnSessionCli.run(FlinkYarnSessionCli.java:717)
    at org.apache.flink.yarn.cli.FlinkYarnSessionCli$1.call(FlinkYarnSessionCli.java:514)
    at org.apache.flink.yarn.cli.FlinkYarnSessionCli$1.call(FlinkYarnSessionCli.java:511)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1729)
    at org.apache.flink.runtime.security.HadoopSecurityContext.runSecured(HadoopSecurityContext.java:41)
    at org.apache.flink.yarn.cli.FlinkYarnSessionCli.main(FlinkYarnSessionCli.java:511)
Caused by: org.apache.flink.runtime.leaderretrieval.LeaderRetrievalException: Could not retrieve the leader
gateway.
    at org.apache.flink.runtime.util.LeaderRetrievalUtils.retrieveLeaderGateway(LeaderRetrievalUtils.java:79)
    at org.apache.flink.client.program.ClusterClient.getJobManagerGateway(ClusterClient.java:856)
    ... 10 common frames omitted
Caused by: java.util.concurrent.TimeoutException: Futures timed out after [10000 milliseconds]
```

Possible Causes

The SSL communication encryption is enabled for Flink, but no correct SSL certificate is configured.

Solution

For MRS 2.*x* or earlier, perform the following operations:

Method 1:

Run the following command to disable the Flink SSL communication encryption, and modify the client configuration file **conf/flink-conf.yaml**.

```
security.ssl.internal.enabled: false
```

Method 2:

Enable the Flink SSL communication encryption and retain the default value of **security.ssl.internal.enabled**. Configure the SSL as follows:

- If the KeyStore or TrustStore file is a relative path, and the Flink client directory where the command is executed can directly access this relative path.

  ```
  security.ssl.internal.keystore: ssl/flink.keystore
  security.ssl.internal.truststore: ssl/flink.truststore
  ```

  Add **-t** option to the CLI **yarn-session.sh** command of Flink to transmit the KeyStore and TrustStore files to each execution node. Example:

  ```
  yarn-session.sh -t ssl/ 2
  ```

- If the keystore or truststore file path is an absolute path, the keystore or truststore files must exist in the absolute path on Flink Client and all nodes.

  ```
  security.ssl.internal.keystore: /opt/client/Flink/flink/conf/flink.keystore
  security.ssl.internal.truststore: /opt/client/Flink/flink/conf/flink.truststore
  ```

For MRS 3.*x* or later, perform the following operations:

Method 1:

Run the following command to disable the Flink SSL communication encryption, and modify the client configuration file **conf/flink-conf.yaml**.

```
security.ssl.enabled: false
```

Method 2:

Enable the Flink SSL communication encryption and retain the default value of **security.ssl.enabled**. Configure the SSL as follows:

- If the KeyStore or TrustStore file is a relative path, and the Flink client directory where the command is executed can directly access this relative path.
```
security.ssl.keystore: ssl/flink.keystore
security.ssl.truststore: ssl/flink.truststore
```
  Add **-t** option to the CLI **yarn-session.sh** command of Flink to transmit the KeyStore and TrustStore files to each execution node. Example:
```
yarn-session.sh -t ssl/ 2
```
- If the keystore or truststore file path is an absolute path, the keystore or truststore files must exist in the absolute path on Flink Client and all nodes.
```
security.ssl.keystore: /opt/Bigdata/client/Flink/flink/conf/flink.keystore
security.ssl.truststore: /opt/Bigdata/client/Flink/flink/conf/flink.truststore
```

15.6.4 Failed to Create a Cluster by Executing the yarn-session Command When a Different User Is Used

Symptom

Two users **testuser** and **bdpuser** with the same rights are used to create the Flink cluster.

When user **testuser** is used to create a Flink cluster, no error message is displayed. While user **bdpuser** is used to create a Flink cluster, an error message is displayed during the **yarn-session.sh** command execution:

```
2019-01-02 14:28:09,098 | ERROR | [main] | Ensure path threw exception |
org.apache.flink.shaded.curator.org.apache.curator.framework.imps.CuratorFrameworkImpl
(CuratorFrameworkImpl.java:566)
org.apache.flink.shaded.zookeeper.org.apache.zookeeper.KeeperException$NoAuthException:
KeeperErrorCode = NoAuth for /flink/application_1545397824912_0022
```

Possible Causes

The HA configuration item is not modified. In the Flink configuration file, the default value of **high-availability.zookeeper.client.acl** is **creator**, indicating that only the creator has the access permission. A new user cannot access the directory on ZooKeeper. As a result, the **yarn-session.sh** command execution fails.

Solution

Step 1 Modify the value of **high-availability.zookeeper.path.root** in the **conf/flink-conf.yaml** file. For example, run the following command:

```
high-availability.zookeeper.path.root: flink2
```

Step 2 Submit the tasks again.

----End

15.6.5 Flink Service Program Fails to Read Files on the NFS Disk

Issue

The Flink service program cannot read files on the NFS disk mounted to the cluster node.

Symptom

The Flink service program developed by a user needs to read the user-defined configuration file. The configuration file is stored on the NFS disk. The NFS disk is mounted to the cluster node and can be accessed by all nodes in the cluster. After the user submits the Flink program, the service code cannot access the user-defined configuration file. As a result, the service program fails to be started.

Cause Analysis

The root cause is that the permission on the root directory of the NFS disk is insufficient. As a result, the Flink program cannot access the directory after being started.

Flink tasks of MRS are running on Yarn. If the cluster is a common cluster, the user who runs the tasks on Yarn is **yarn_user**. If the user-defined configuration file is used after the tasks are started, **yarn_user** must be allowed to access the file and the parent directory of the file (parent directory of the file on the NFS, not the soft link on the cluster node). Otherwise, the program cannot obtain the file content. If the cluster is a cluster with Kerberos authentication enabled, the file permission must allow the user who submits the program to access the file.

Procedure

Step 1 Log in to the Master node in the cluster as user **root**.

Step 2 Run the following command to check the permission on the parent directory of the user-defined configuration file:

ll <*Parent directory of the file path***>**

Step 3 Go to the directory of the file to be accessed on the NFS disk and change the permission of the parent directory of the user-defined configuration file to 755.

chmod 755 -R /<*Path of the parent directory of the file***>**

Step 4 Check whether the Core or Task node can access the configuration file.

1. Log in to the Core or Task node as the **root** user.

   If Kerberos authentication is enabled for the current cluster, log in to the Core node as user **root**.

2. Run **su - yarn_user** to switch to user **yarn_user**.

   If Kerberos authentication is enabled for the cluster, run the **su -** *User who submits the job* command to switch the user.

3. Run the following command to check the user permission. The file path must be the absolute path of the file.

ll <*File path*>

----End

Summary and Suggestions

When a user-defined configuration file needs to be accessed in the submitted task, especially when the NFS disk is mounted, you need to check whether the permission of the parent directory of the file is correct in addition to the file permission. When an NFS disk is mounted to an MRS cluster node, a soft link is created to the NFS directory. In this case, you need to check whether the directory permission on the NFS is correct.

15.6.6 Failed to Customize the Flink Log4j Log Level

Issue

The customized level for Flink Log4j logs of an MRS 3.1.0 cluster does not take effect.

Symptom

1. When analyzing data using Flink of an MRS 3.1.0 cluster, a user changes the log level in the **log4j.properties** file in the **$Flink_HOME/conf** directory to **INFO**.
2. However, after the task is submitted successfully, the log level displayed on the console is still **ERROR**, rather than **INFO**.

Cause Analysis

The **log4j.properties** file in the **$Flink_HOME/conf** directory controls the log output of in JobManager and TaskManager operators, and the logs are printed to the corresponding Yarn containers. You can view the logs on the Yarn web UI. In MRS 3.1.0 and later versions, the default log framework of Flink 1.12.0 is Log4j2. The configuration method is different from that of Log4j. For example, Log4j log rules do not take effect.

Procedure

For details about configuring Log4j2 log specifications, see the official open-source document at **http://logging.apache.org/log4j/2.x/manual/configuration.html#Properties**.

15.7 Using Flume

15.7.1 Class Cannot Be Found After Flume Submits Jobs to Spark Streaming

Issue

After Flume submits jobs to Spark Streaming, the class cannot be found.

Symptom

After the Spark Streaming code is packed into a JAR file and submitted to the cluster, an error message is displayed indicating that the class cannot be found. The following two methods are not useful:

1. When submitting a Spark job, run the **--jars** command to reference the JAR file of the class.
2. Import the JAR file where the class resides to the JAR file of Spark Streaming.

Cause Analysis

Some JAR files cannot be loaded during Spark job execution, resulting that the class cannot be found.

Procedure

Step 1 Run the **--jars** command to load the **flume-ng-sdk-{version} .jar** dependency package.

Step 2 Modify the two configuration items in the **spark-default.conf** file:

spark.driver.extraClassPath=$PWD/*: {Add the original value}

spark.executor.extraClassPath =$PWD/*

Step 3 Run the job successfully. If an error is reported, check which JAR is not loaded and perform step 1 and step 2 again.

----End

15.7.2 Failed to Install a Flume Client

Symptom

A Flume client fails to be installed, and "JAVA_HOME is null" or "flume has been installed" is displayed.

```
CST 2016-08-31 17:02:51 [flume-client install]: JAVA_HOME is null in current user,please install the JDK and set the JAVA_HOME
CST 2016-08-31 17:02:51 [flume-client install]: check environment failed.
CST 2016-08-31 17:02:51 [flume-client install]: check param failed.
CST 2016-08-31 17:02:51 [flume-client install]: install flume client failed.
```

```
CST 2016-08-31 17:03:58 [flume-client install]: flume has been installed
CST 2016-08-31 17:03:58 [flume-client install]: check path failed.
CST 2016-08-31 17:03:58 [flume-client install]: check param failed.
CST 2016-08-31 17:03:58 [flume-client install]: install flume client failed.
```

Cause Analysis

- Environment variables are checked during Flume client installation. If no Java is available, an error message is displayed stating "JAVA_HOME is null" and the installation quits.
- If Flume has been installed in the specified directory, an error message is displayed stating "flume has been installed" during client installation and the installation quits.

Solution

Step 1 Run the following command if an error message is displayed stating "JAVA_HOME is null":

export JAVA_HOME=*Java path*

Set **JAVA_HOME** and execute the installation script again.

Step 2 If a Flume client has been installed under the specified directory, uninstall the client and use another directory.

----End

15.7.3 A Flume Client Cannot Connect to the Server

Symptom

A user installs a Flume client and sets an Avro sink to communicate with the server. However, the Flume server cannot be connected.

Cause Analysis

1. The server is incorrectly configured and the monitoring port fails to be started up. For example, an incorrect IP address or an occupied port is configured for the Avro source of the server. View Flume run logs.

```
2016-08-31 17:28:42,092 | ERROR | [lifecycleSupervisor-1-9] |  Unable to start
EventDrivenSourceRunner: { source:Avro source avro_source: { bindAddress: 10.120.205.7, port:
21154 } } - Exception follows.  | org.apache.flume.lifecycle.LifecycleSupervisor
$MonitorRunnable.run(LifecycleSupervisor.java:253)
java.lang.RuntimeException: org.jboss.netty.channel.ChannelException: Failed to bind to: /
192.168.205.7:21154
```

2. If encrypted transmission is used, the certificate or password is incorrect.

```
2016-08-31 17:15:59,593 | ERROR | [conf-file-poller-0] |  Source avro_source has been removed due to
an error during configuration  |
org.apache.flume.node.AbstractConfigurationProvider.loadSources(AbstractConfigurationProvider.java:
388)
org.apache.flume.FlumeException: Avro source configured with invalid keystore: /opt/Bigdata/
MRS_XXX/install/FusionInsight-Flume-1.9.0/flume/conf/flume_sChat.jks
```

3. The network connection between the client and the server is abnormal.

```
PING 192.168.85.55 (10.120.85.55) 56(84) bytes of data.
From 192.168.85.50 icmp_seq=1 Destination Host Unreachable
From 192.168.85.50 icmp_seq=2 Destination Host Unreachable
From 192.168.85.50 icmp_seq=3 Destination Host Unreachable
From 192.168.85.50 icmp_seq=4 Destination Host Unreachable
```

Solution

Step 1 Set a correct IP address (an IP address of the local host). If the port has been occupied, configure another free port.

Step 2 Configure a correct certificate path.

Step 3 Contact the network administrator to restore the network.

----End

15.7.4 Flume Data Fails to Be Written to the Component

Symptom

After the Flume process is started, Flume data cannot be written to the corresponding component. (The following uses writing data from the server to HDFS as an example.)

Cause Analysis

1. HDFS is not started or is faulty. View Flume run logs.

```
2019-02-26 11:16:33,564 | ERROR | [SinkRunner-PollingRunner-DefaultSinkProcessor] |  opreation the
hdfs file errors.  | org.apache.flume.sink.hdfs.HDFSEventSink.process(HDFSEventSink.java:414)
2019-02-26 11:16:33,747 | WARN  | [hdfs-CCCC-call-runner-4] |  A failover has occurred since the start
of call #32795 ClientNamenodeProtocolTranslatorPB.getFileInfo over
192-168-13-88/192.168.13.88:25000  | org.apache.hadoop.io.retry.RetryInvocationHandler
$ProxyDescriptor.failover(RetryInvocationHandler.java:220)
2019-02-26 11:16:33,748 | ERROR | [hdfs-CCCC-call-runner-4] |  execute hdfs error. {}  |
org.apache.flume.sink.hdfs.HDFSEventSink$3.call(HDFSEventSink.java:744)
java.net.ConnectException: Call From 192-168-12-221/192.168.12.221 to 192-168-13-88:25000 failed
on connection exception: java.net.ConnectException: Connection refused; For more details see:  http://
wiki.apache.org/hadoop/ConnectionRefused
```

2. The HDFS sink is not started. Check the Flume run log. It is found that the Flume current metrics file does not contain sink information.

```
2019-02-26 11:46:05,501 | INFO  | [pool-22-thread-1] |  flume current metrics:{"CHANNEL.BBBB":
{"ChannelCapacity":"10000","ChannelFillPercentage":"0.0","Type":"CHANNEL","ChannelStoreSize":"0","
EventProcessTimedelta":"0","EventTakeSuccessCount":"0","ChannelSize":"0","EventTakeAttemptCount":
"0","StartTime":"1551152734999","EventPutAttemptCount":"0","EventPutSuccessCount":"0","StopTime
":"0"},"SOURCE.AAAA":
{"AppendBatchAcceptedCount":"0","EventAcceptedCount":"0","AppendReceivedCount":"0","MonTime":
"0","StartTime":"1551152735503","AppendBatchReceivedCount":"0","EventReceivedCount":"0","Type":"
SOURCE","TotalFilesCount":"1001","SizeAcceptedCount":"0","UpdateTime":"605410241202740","Appen
dAcceptedCount":"0","OpenConnectionCount":"0","MovedFilesCount":"1001","StopTime":"0"}}  |
org.apache.flume.node.Application.getRestartComps(Application.java:467)
```

Solution

Step 1 If the component to which Flume writes data is not started, start the component. If the component is abnormal, contact technical support.

Step 2 If the sink is not started, check whether the configuration file is correctly configured. If the configuration file is incorrectly configured, modify the configuration file and restart the Flume process. If the configuration file is correctly configured, view the error information in the log and rectify the fault based on the error information.

----End

15.7.5 Flume Server Process Fault

Symptom

After Flume runs for a period of time, the Flume instance is in the faulty state on Manager.

Cause Analysis

If the Flume file or folder permission is abnormal, the following information is displayed on MRS Manager after the restart:

```
[2019-02-26 13:38:02]RoleInstance prepare to start failure [{ScriptExecutionResult=ScriptExecutionResult
[exitCode=126, output=, errMsg=sh: line 1: /opt/Bigdata/MRS_XXX/install/FusionInsight-Flume-1.9.0/
flume/bin/flume-manage.sh: Permission denied
```

Solution

Compare the file and folder permissions with those for the Flume node that is running properly and correct the file or folder permissions.

15.7.6 Flume Data Collection Is Slow

Symptom

After Flume is started, it takes a long time for Flume to collect data.

Cause Analysis

1. The heap memory of Flume is not properly set. As a result, the Flume process keeps in the GC state. View Flume run logs.
   ```
   2019-02-26T13:06:20.666+0800: 1085673.512: [Full GC:[CMS: 3849339k->3843458K(3853568K),
   2.5817610 secs] 4153654K->3843458K(4160256K), [CMS Perm : 27335K->27335K(45592K),2.5820080
   SECS] [Times: user=2.63, sys0.00, real=2.59 secs]
   ```
2. The **deletePolicy** policy configured for the Spooldir source is **immediate**.

Solution

Step 1 Increase the size of the heap memory (**xmx**).

Step 2 Change the **deletePolicy** policy of the Spooldir source to **never**.

----End

15.7.7 Failed to Start Flume

Symptom

The Flume service fails to be installed or restarted.

Cause Analysis

1. The heap memory of Flume is greater than the remaining memory of the server. The Flume startup log shows the following information:

```
[CST 2019-02-26 13:31:43][INFO] [[checkMemoryValidity:124]] [GC_OPTS is invalid:
Xmx(40960000MB) is bigger than the free memory(56118MB) in system.] [9928]
```

2. The permission on the Flume file or folder is abnormal. The following information is displayed on the GUI or in the background:

```
[2019-02-26 13:38:02]RoleInstance prepare to start failure
[{ScriptExecutionResult=ScriptExecutionResult [exitCode=126, output=, errMsg=sh: line 1: /opt/Bigdata/
MRS_XXX/install/FusionInsight-Flume-1.9.0/flume/bin/flume-manage.sh: Permission denied
```

3. The **JAVA_HOME** is incorrectly configured. The Flume agent startup log shows the following information:

```
Info: Sourcing environment configuration script /opt/FlumeClient/fusioninsight-flume-1.9.0/conf/
flume-env.sh
+ '[' -n '' ']'
+ exec /tmp/MRS-Client/MRS_Flume_ClientConfig/JDK/jdk-8u18/bin/java '-
XX:OnOutOfMemoryError=bash /opt/FlumeClient/fusioninsight-flume-1.9.0/bin/
out_memory_error.sh /opt/FlumeClient/fusioninsight-flume-1.9.0/conf %p' -Xms2G -Xmx4G -
XX:CMSFullGCsBeforeCompaction=1 -XX:+UseConcMarkSweepGC -XX:+CMSParallelRemarkEnabled -XX:
+UseCMSCompactAtFullCollection -Dkerberos.domain.name=hadoop.hadoop.com -verbose:gc -XX:
+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=1M -XX:+PrintGCDetails -XX:
+PrintGCDateStamps -Xloggc:/var/log/Bigdata//flume-client-1/flume/flume-root-20190226134231-
%p-gc.log -Dproc_org.apache.flume.node.Application -Dproc_name=client -Dproc_conf_file=/opt/
FlumeClient/fusioninsight-flume-1.9.0/conf/properties.properties -Djava.security.krb5.conf=/opt/
FlumeClient/fusioninsight-flume-1.9.0/conf//krb5.conf -Djava.security.auth.login.config=/opt/
FlumeClient/fusioninsight-flume-1.9.0/conf//jaas.conf -Dzookeeper.server.principal=zookeeper/
hadoop.hadoop.com -Dzookeeper.request.timeout=120000 -Dflume.instance.id=884174180 -
Dflume.agent.name=clientName1 -Dflume.role=client -Dlog4j.configuration.watch=true -
Dlog4j.configuration=log4j.properties -Dflume_log_dir=/var/log/Bigdata//flume-client-1/flume/ -
Dflume.service.id=flume-client-1 -Dbeetle.application.home.path=/opt/FlumeClient/fusioninsight-
flume-1.9.0/conf/service -Dflume.called.from.service -Dflume.conf.dir=/opt/FlumeClient/fusioninsight-
flume-1.9.0/conf -Dflume.metric.conf.dir=/opt/FlumeClient/fusioninsight-flume-1.9.0/conf -
Dflume.script.home=/opt/FlumeClient/fusioninsight-flume-1.9.0/bin -cp '/opt/FlumeClient/
fusioninsight-flume-1.9.0/conf:/opt/FlumeClient/fusioninsight-flume-1.9.0/lib/*:/opt/FlumeClient/
fusioninsight-flume-1.9.0/conf/service/' -Djava.library.path=/opt/FlumeClient/fusioninsight-flume-1.9.0/
plugins.d/native/native org.apache.flume.node.Application --conf-file /opt/FlumeClient/fusioninsight-
flume-1.9.0/conf/properties.properties --name client
/opt/FlumeClient/fusioninsight-flume-1.9.0/bin/flume-ng: line 233: /tmp/FusionInsight-Client/Flume/
FusionInsight_Flume_ClientConfig/JDK/jdk-8u18/bin/java: No such file or directory
```

Solution

Step 1 Increase the size of the heap memory (**xmx**).

Step 2 Compare the file and folder permissions with those for node where Flume is started properly and change the incorrect file or folder permissions.

Step 3 Reconfigure **JAVA_HOME**. On the client, replace the value of **JAVA_HOME** in the **${install_home}/fusioninsight-flume-***Flume version***/conf/ENV_VARS** file. On the server, replace the value of **JAVA_HOME** in the **ENV_VARS** file in the **etc** directory.

To obtain the value of **JAVA_HOME**, log in to the node where Flume is properly started and run the **echo ${JAVA_HOME}** command.

NOTE

${install_home} is the installation path of the Flume client.

----End

15.8 Using HBase

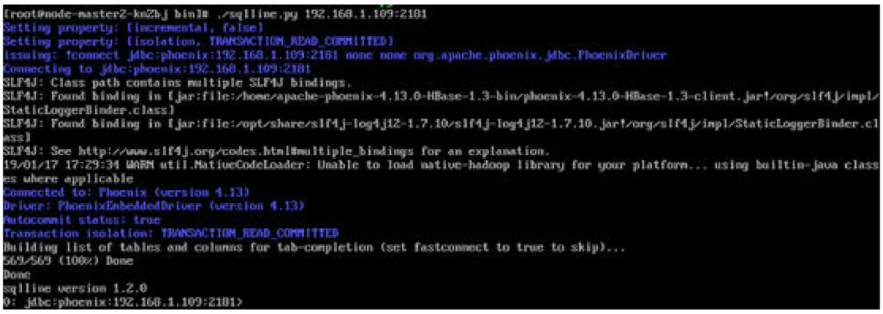
15.8.1 Slow Response to HBase Connection

Issue

Under the same VPC network, response is slow when an external cluster connects to HBase through Phoenix.

Symptom

Under the same VPC network, response is slow when an external cluster connects to HBase through Phoenix.



```
[root@node-master2-knZbj bin]# ./sqlline.py 192.168.1.109:2181
Setting property: [incremental, false]
Setting property: [isolation, TRANSACTION_READ_COMMITTED]
issuing: !connect jdbc:phoenix:192.168.1.109:2181 none none org.apache.phoenix.jdbc.PhoenixDriver
Connecting to jdbc:phoenix:192.168.1.109:2181
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/apache-phoenix-4.13.0-HBase-1.3-bin/phoenix-4.13.0-HBase-1.3-client.jar!/org/slf4j/impl/
StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/share/slf4j-log4j12-1.7.10/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.cl
ass]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
19/01/17 17:29:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java class
es where applicable
Connected to: Phoenix (version 4.13)
Driver: PhoenixEmbeddedDriver (version 4.13)
Autocommit status: true
Transaction isolation: TRANSACTION_READ_COMMITTED
Building list of tables and columns for tab-completion (set fastconnect to true to skip)...
569/569 (100%) Done
Done
sqlline version 1.2.0
0: jdbc:phoenix:192.168.1.109:2181>
```

Possible Cause

DNS has been configured. When a client connects to HBase, DNS resolves the server first, causing slow response.

Procedure

Step 1 Log in to the Master node as user **root**.

Step 2 Run the **vi /etc/resolv.conf** command to open the **resolv.conf** file and comment out the address of the DNS server, for example, #1.1.1.1.

----End

15.8.2 Failed to Authenticate the HBase User

Issue

Failed to authenticate the HBase user.

Symptom

Failed to authenticate the HBase user on the client. The following error information is displayed:

```
2019-05-13 10:53:09,975 ERROR [localhost-startStop-1] xxxConfig.LoginUtil: login failed with hbaseuser
and /usr/local/linoseyc/hbase-tomcat/webapps/bigdata_hbase/WEB-INF/classes/user.keytab.
2019-05-13 10:53:09,975 ERROR [localhost-startStop-1] xxxConfig.LoginUtil: perhaps cause 1 is (wrong
password) keytab file and user not match, you can kinit -k -t keytab user in client server to check.
2019-05-13 10:53:09,975 ERROR [localhost-startStop-1] xxxConfig.LoginUtil: perhaps cause 2 is (clock
skew) time of local server and remote server not match, please check ntp to remote server.
2019-05-13 10:53:09,975 ERROR [localhost-startStop-1] xxxConfig.LoginUtil: perhaps cause 3 is (aes256 not
support) aes256 not support by default jdk/jre, need copy local_policy.jar and US_export_policy.jar from
remote server in path ${BIGDATA_HOME}/jdk/jre/lib/security.
```

Cause Analysis

The version of the JAR file in the JDK used by the user is different from that of the JAR file authenticated by MRS.

Procedure

Step 1 Log in to the Master1 node as user **root**.

Step 2 Run the following command to check the JAR file authenticated by MRS:

ll /opt/share/local_policy/local_policy.jar

ll /opt/Bigdata/jdk{version}/jre/lib/security/local_policy.jar

Step 3 Download the JAR package queried in step 2 to the local host.

Step 4 Copy the downloaded JAR package to the local JDK directory **/opt/Bigdata/jdk/jre/lib/security**.

Step 5 Run the **cd /opt/client/HBase/hbase/bin** command to go to the **bin** directory of HBase.

Step 6 Run the **sh start-hbase.sh** command to restart HBase.

----End

15.8.3 RegionServer Failed to Start Because the Port Is Occupied

Symptom

RegionServer is in the **Restoring** state on Manager.

Cause Analysis

1. View the RegionServer log (**/var/log/Bigdata/hbase/rs/hbase-omm-xxx.log**).
2. Run the **lsof -i:21302** command (the port number of MRS 1.7.X and later versions is 16020) to view the PID. Based on the PID, check the process. It is found that the RegionServer port is occupied by DFSZkFailoverController.
3. The value of **/proc/sys/net/ipv4/ip_local_port_range** is **9000 65500**. The temporary port range and the MRS port range overlap. This is because the preinstall operation is not performed during installation.

Solution

Step 1 Run the **kill -9** *DFSZkFailoverController pid* command to ensure that another port is bound with after a restart and restart the RegionServer in the **Restoring** state.

----End

15.8.4 HBase Failed to Start Due to Insufficient Node Memory

Symptom

The RegionServer service of HBase is always in the **Restoring** state.

Cause Analysis

1. Check the RegionServer log (**/var/log/Bigdata/hbase/rs/hbase-omm-XXX.out**). It is found that the following information is printed:

   ```
   There is insufficient memory for the Java Runtime Environment to continue.
   ```

2. Run the **free** command to check the memory. It is found that the available memory of the node is insufficient.

Solution

Step 1 Locate why the memory is insufficient. It is found that some processes occupy too much memory or the server does not have sufficient memory.

----End

15.8.5 HBase Service Unavailable Due to Poor HDFS Performance

Symptom

The HBase component intermittently reports alarms indicating that the service is unavailable.

Cause Analysis

HDFS performance is low, causing health check timeout and the alarm is generated accordingly. You can perform the following operations:

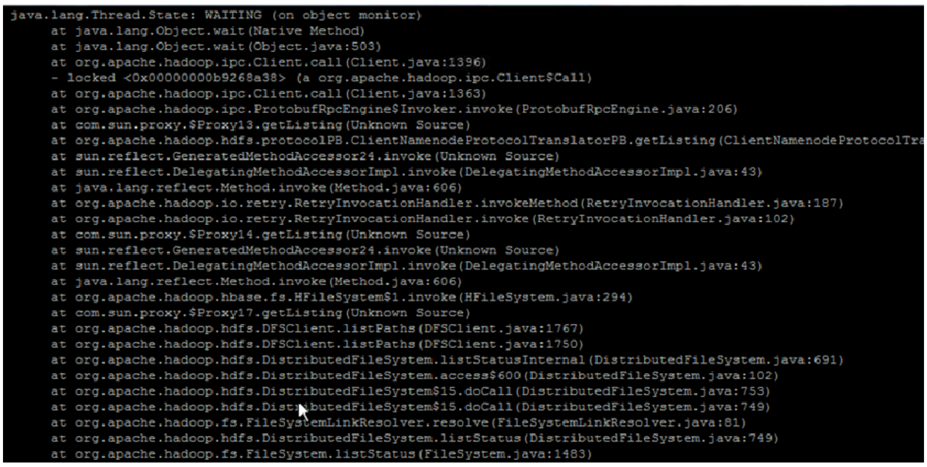
1. View the HMaster log (**/var/log/Bigdata/hbase/hm/hbase-omm-xxx.log**) and check that **system pause**, **jvm**, and other GC-related information is not frequently printed in the log.
2. Determine whether the fault is caused by poor HDFS performance using either of the following methods:
   a. Run **hbase shell** to access the HBase shell, and run the **list** command to check whether it takes a long period of time to list all tables in HBase.
   b. Enable printing of the debug logs of HDFS, and check whether it takes a long period of time to list the content of a large number of directories by running the **hadoop fs –ls** */XXX/XXX* command.
   c. Print the Java stack information about a specified HMaster process.

      su - omm

      jps

      jstack pid
3. Check the jstack information. The following figure shows that the process is stuck at the **DFSClient.listPaths** state.

Figure 15-20 Exception



```
java.lang.Thread.State: WAITING (on object monitor)
      at java.lang.Object.wait(Native Method)
      at java.lang.Object.wait(Object.java:503)
      at org.apache.hadoop.ipc.Client.call(Client.java:1396)
      - locked <0x00000000b9268a38> (a org.apache.hadoop.ipc.Client$Call)
      at org.apache.hadoop.ipc.Client.call(Client.java:1363)
      at org.apache.hadoop.ipc.ProtobufRpcEngine$Invoker.invoke(ProtobufRpcEngine.java:206)
      at com.sun.proxy.$Proxy13.getListing(Unknown Source)
      at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.getListing(ClientNamenodeProtocolTra
      at sun.reflect.GeneratedMethodAccessor24.invoke(Unknown Source)
      at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
      at java.lang.reflect.Method.invoke(Method.java:606)
      at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:187)
      at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:102)
      at com.sun.proxy.$Proxy14.getListing(Unknown Source)
      at sun.reflect.GeneratedMethodAccessor24.invoke(Unknown Source)
      at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
      at java.lang.reflect.Method.invoke(Method.java:606)
      at org.apache.hadoop.hbase.fs.HFileSystem$1.invoke(HFileSystem.java:294)
      at com.sun.proxy.$Proxy17.getListing(Unknown Source)
      at org.apache.hadoop.hdfs.DFSClient.listPaths(DFSClient.java:1767)
      at org.apache.hadoop.hdfs.DFSClient.listPaths(DFSClient.java:1750)
      at org.apache.hadoop.hdfs.DistributedFileSystem.listStatusInternal(DistributedFileSystem.java:691)
      at org.apache.hadoop.hdfs.DistributedFileSystem.access$600(DistributedFileSystem.java:102)
      at org.apache.hadoop.hdfs.DistributedFileSystem$15.doCall(DistributedFileSystem.java:753)
      at org.apache.hadoop.hdfs.DistributedFileSystem$15.doCall(DistributedFileSystem.java:749)
      at org.apache.hadoop.fs.FileSystemLinkResolver.resolve(FileSystemLinkResolver.java:81)
      at org.apache.hadoop.hdfs.DistributedFileSystem.listStatus(DistributedFileSystem.java:749)
      at org.apache.hadoop.fs.FileSystem.listStatus(FileSystem.java:1483)
```

Solution

Step 1 If this alarm is caused by poor HDFS performance, check whether Impala is of an earlier version or JournalNode was incorrectly deployed during the initial deployment (more than three JournalNode nodes are deployed).

----End

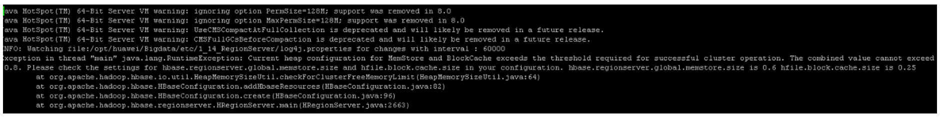
15.8.6 HBase Failed to Start Due to Inappropriate Parameter Settings

Symptom

After some parameters are modified, HBase cannot be started.

Cause Analysis

1. Check the HMaster log (**/var/log/Bigdata/hbase/hm/hbase-omm-*xxx*.log**). It is found that the total of **hbase.regionserver.global.memstore.size** and **hfile.block.cache.size** is greater than 0.8, which causes the startup failure. Therefore, adjust the parameter values to make sure that the total value is less than 0.8.



```
ava HotSpot(TM) 64-Bit Server VM warning: ignoring option PermSize=128M; support was removed in 8.0
ava HotSpot(TM) 64-Bit Server VM warning: ignoring option MaxPermSize=128M; support was removed in 8.0
ava HotSpot(TM) 64-Bit Server VM warning: UseCMSCompactAtFullCollection is deprecated and will likely be removed in a future release.
ava HotSpot(TM) 64-Bit Server VM warning: CMSFullGCsBeforeCompaction is deprecated and will likely be removed in a future release.
NFO: Watching file:/opt/huawei/Bigdata/etc/1_14_RegionServer/log4j.properties for changes with interval : 60000
xception in thread "main" java.lang.RuntimeException: Current heap configuration for MemStore and BlockCache exceeds the threshold required for successful cluster operation. The combined value cannot exceed
0.8. Please check the settings for hbase.regionserver.global.memstore.size and hfile.block.cache.size in your configuration. hbase.regionserver.global.memstore.size is 0.6 hfile.block.cache.size is 0.25
        at org.apache.hadoop.hbase.io.util.HeapMemorySizeUtil.checkForClusterFreeMemoryLimit(HeapMemorySizeUtil.java:64)
        at org.apache.hadoop.hbase.HBaseConfiguration.addHbaseResources(HBaseConfiguration.java:82)
        at org.apache.hadoop.hbase.HBaseConfiguration.create(HBaseConfiguration.java:96)
        at org.apache.hadoop.hbase.regionserver.HRegionServer.main(HRegionServer.java:2663)
```

2. Check the HMaster and RegionServer out logs (**/var/log/Bigdata/hbase/hm/hbase-omm-*xxx*.out/var/log/Bigdata/hbase/rs/hbase-omm-*xxx*.out**). It is found that **Unrecognized VM option** is displayed.

```
Unrecognized VM option
Error: Could not create the Java Virtual Machine.
Error: A fatal exception has occurred. Program will exit.
```

Check the **GC_OPTS** parameters. It is found that the parameters contain unnecessary spaces, for example, **-D sun.rmi.dgc.server.gcInterval=0x7FFFFFFFFFFFFFE**.

Solution

Step 1 After the **MemStore** and **cache** parameters are modified, the HBase service is restarted successfully.

Step 2 After the **GC_OPTS** parameters are modified, the HBase service is restarted successfully.

----End

15.8.7 RegionServer Failed to Start Due to Residual Processes

Symptom

The HBase service fails to start, and an error is reported during the health check.

Cause Analysis

Check detailed information about HBase startup on the MRS Manager page. It is found that **the previous process is not quit** is displayed.

Solution

Step 1 Log in to the node and run the **ps -ef | grep HRegionServer** command in the background. A residual process is found.

Step 2 After confirming that the process can be killed, kill the process. If the process cannot be stopped by running the **kill** command, run the **kill -9** command to forcibly stop the process.

Step 3 Restart the HBase service.

----End

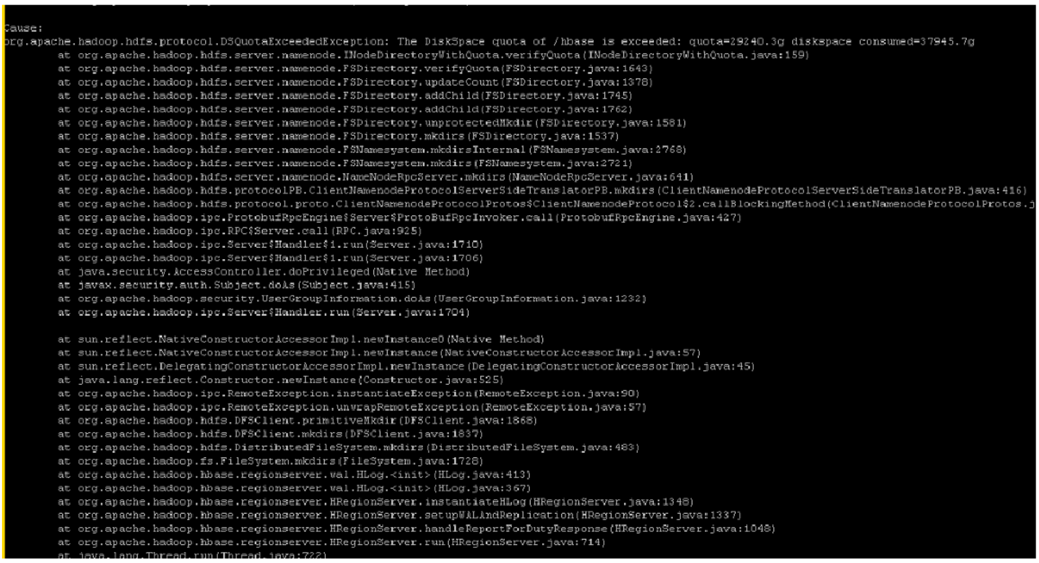
15.8.8 HBase Failed to Start Due to a Quota Set on HDFS

Symptom

HBase fails to start.

Cause Analysis

Check the HMaster log (**/var/log/Bigdata/hbase/hm/hbase-omm-*xxx*.log**). It is found that "The DiskSpace quota of /hbase is exceeded" is displayed.



```
Cause:
org.apache.hadoop.hdfs.protocol.DSQuotaExceededException: The DiskSpace quota of /hbase is exceeded: quota=29240.3g diskspace consumed=37945.7g
        at org.apache.hadoop.hdfs.server.namenode.INodeDirectoryWithQuota.verifyQuota(INodeDirectoryWithQuota.java:159)
        at org.apache.hadoop.hdfs.server.namenode.FSDirectory.verifyQuota(FSDirectory.java:1643)
        at org.apache.hadoop.hdfs.server.namenode.FSDirectory.updateCount(FSDirectory.java:1378)
        at org.apache.hadoop.hdfs.server.namenode.FSDirectory.addChild(FSDirectory.java:1745)
        at org.apache.hadoop.hdfs.server.namenode.FSDirectory.addChild(FSDirectory.java:1762)
        at org.apache.hadoop.hdfs.server.namenode.FSDirectory.unprotectedMkdir(FSDirectory.java:1581)
        at org.apache.hadoop.hdfs.server.namenode.FSDirectory.mkdirs(FSDirectory.java:1537)
        at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirsInternal(FSNamesystem.java:2768)
        at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirs(FSNamesystem.java:2721)
        at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.mkdirs(NameNodeRpcServer.java:641)
        at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolServerSideTranslatorPB.mkdirs(ClientNamenodeProtocolServerSideTranslatorPB.java:416)
        at org.apache.hadoop.hdfs.protocol.proto.ClientNamenodeProtocolProtos$ClientNamenodeProtocol$2.callBlockingMethod(ClientNamenodeProtocolProtos.j
        at org.apache.hadoop.ipc.ProtobufRpcEngine$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine.java:427)
        at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:925)
        at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:1710)
        at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:1706)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:415)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1232)
        at org.apache.hadoop.ipc.Server$Handler.run(Server.java:1704)

        at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native Method)
        at sun.reflect.NativeConstructorAccessorImpl.newInstance(NativeConstructorAccessorImpl.java:57)
        at sun.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:45)
        at java.lang.reflect.Constructor.newInstance(Constructor.java:525)
        at org.apache.hadoop.ipc.RemoteException.instantiateException(RemoteException.java:90)
        at org.apache.hadoop.ipc.RemoteException.unwrapRemoteException(RemoteException.java:57)
        at org.apache.hadoop.hdfs.DFSClient.primitiveMkdir(DFSClient.java:1868)
        at org.apache.hadoop.hdfs.DFSClient.mkdirs(DFSClient.java:1837)
        at org.apache.hadoop.hdfs.DistributedFileSystem.mkdirs(DistributedFileSystem.java:483)
        at org.apache.hadoop.fs.FileSystem.mkdirs(FileSystem.java:1728)
        at org.apache.hadoop.hbase.regionserver.wal.HLog.<init>(HLog.java:413)
        at org.apache.hadoop.hbase.regionserver.wal.HLog.<init>(HLog.java:367)
        at org.apache.hadoop.hbase.regionserver.HRegionServer.instantiateHLog(HRegionServer.java:1348)
        at org.apache.hadoop.hbase.regionserver.HRegionServer.setupWALAndReplication(HRegionServer.java:1337)
        at org.apache.hadoop.hbase.regionserver.HRegionServer.handleReportForDutyResponse(HRegionServer.java:1048)
        at org.apache.hadoop.hbase.regionserver.HRegionServer.run(HRegionServer.java:714)
        at java.lang.Thread.run(Thread.java:722)
```

Solution

Step 1 Run the **df -h** command to check data directory space. It is found that the directory space is full. Delete unnecessary data to free up space.

Step 2 Expand the node to ensure that the data directory space is sufficient.

----End

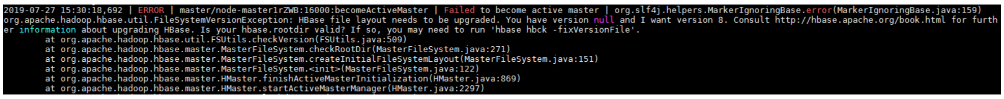
15.8.9 HBase Failed to Start Due to Corrupted Version Files

Symptom

HBase fails to start.

Cause Analysis

1. The **hbase.version** file is read during HBase startup. However, the log indicates that a reading exception occurs.



```
2019-07-27 15:30:18,692 | ERROR | master/node-master1rZWB:16000:becomeActiveMaster | Failed to become active master | org.slf4j.helpers.MarkerIgnoringBase.error(MarkerIgnoringBase.java:159)
org.apache.hadoop.hbase.util.FileSystemVersionException: HBase file layout needs to be upgraded. You have version null and I want version 8. Consult http://hbase.apache.org/book.html for furth
er information about upgrading HBase. Is your hbase.rootdir valid? If so, you may need to run 'hbase hbck -fixVersionFile'.
        at org.apache.hadoop.hbase.util.FSUtils.checkVersion(FSUtils.java:509)
        at org.apache.hadoop.hbase.master.MasterFileSystem.checkRootDir(MasterFileSystem.java:271)
        at org.apache.hadoop.hbase.master.MasterFileSystem.createInitialFileSystemLayout(MasterFileSystem.java:151)
        at org.apache.hadoop.hbase.master.MasterFileSystem.<init>(MasterFileSystem.java:122)
        at org.apache.hadoop.hbase.master.HMaster.finishActiveMasterInitialization(HMaster.java:869)
        at org.apache.hadoop.hbase.master.HMaster.startActiveMasterManager(HMaster.java:2297)
```

2. The file cannot be viewed by running the **hadoop fs -cat /hbase/hbase.version** command. The file is corrupted.

Solution

Step 1 Run the **hbase hbck -fixVersionFile** command to restore the file.

Step 2 If the problem persists after performing **Step 1**, obtain the **hbase.version** file from another cluster of the same version and upload the file to replace the original one.

Step 3 Restart the HBase service.

----End

15.8.10 High CPU Usage Caused by Zero-Loaded RegionServer

Symptom

The CPU usage of RegionServer is high, but there is no service running on RegionServer.

Cause Analysis

1. Run the **top** command to obtain the CPU usage of RegionServer processes and check the IDs of processes with high CPU usage.
2. Obtain the CPU usage of threads under these processes based on the RegionServer process IDs.

   Run the **top -H -p** *<PID>* (replace it with the actual RegionServer process ID). As shown in the following figure, the CPU usage of some threads reaches 80%.

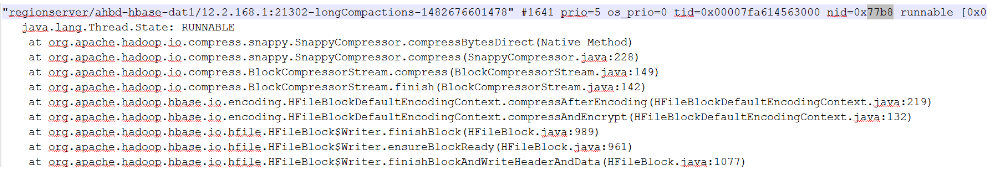
```
PID USER      PR  NI    VIRT    RES    SHR S %CPU %MEM     TIME+ COMMAND
75706 omm      20   0 6879444  1.0g  25612 S  90.4  1.6   0:00.00 java
75716 omm      20   0 6879444  1.0g  25612 S  90.4  1.6   0:04.74 java
75720 omm      20   0 6879444  1.0g  25612 S  88.6  1.6   0:01.93 java
75721 omm      20   0 6879444  1.0g  25612 S  86.8  1.6   0:01.99 java
75722 omm      20   0 6879444  1.0g  25612 S  86.8  1.6   0:01.94 java
75723 omm      20   0 6879444  1.0g  25612 S  86.8  1.6   0:01.96 java
75724 omm      20   0 6879444  1.0g  25612 S  86.8  1.6   0:01.97 java
75725 omm      20   0 6879444  1.0g  25612 S  81.5  1.6   0:02.06 java
75726 omm      20   0 6879444  1.0g  25612 S  79.7  1.6   0:02.01 java
75727 omm      20   0 6879444  1.0g  25612 S  79.7  1.6   0:01.95 java
75728 omm      20   0 6879444  1.0g  25612 S  78.0  1.6   0:01.99 java
```

3. Obtain the thread stack information based on the ID of the RegionServer process.

   jstack 12345 >allstack.txt (Replace it with the actual RegionServer process ID.)
4. Convert the thread ID into the hexadecimal format:

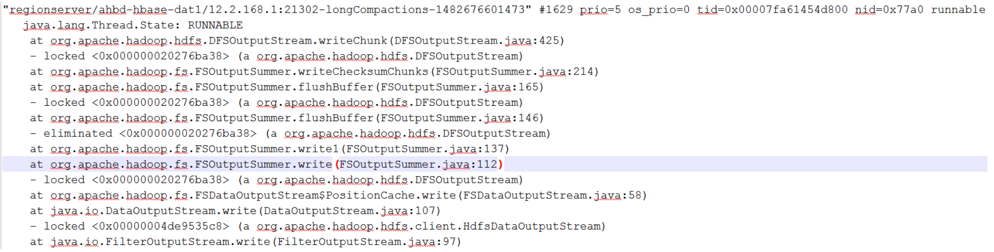
   printf "%x\n" 30648

   In the command output, the TID is **77b8**.
5. Search the thread stack based on the hexadecimal TID. It is found that the compaction operation is performed.



```
"regionserver/ahbd-hbase-dat1/12.2.168.1:21302-longCompactions-1482676601478" #1641 prio=5 os_prio=0 tid=0x00007fa614563000 nid=0x77b8 runnable [0x0
   java.lang.Thread.State: RUNNABLE
    at org.apache.hadoop.io.compress.snappy.SnappyCompressor.compressBytesDirect(Native Method)
    at org.apache.hadoop.io.compress.snappy.SnappyCompressor.compress(SnappyCompressor.java:228)
    at org.apache.hadoop.io.compress.BlockCompressorStream.compress(BlockCompressorStream.java:149)
    at org.apache.hadoop.io.compress.BlockCompressorStream.finish(BlockCompressorStream.java:142)
    at org.apache.hadoop.hbase.io.encoding.HFileBlockDefaultEncodingContext.compressAfterEncoding(HFileBlockDefaultEncodingContext.java:219)
    at org.apache.hadoop.hbase.io.encoding.HFileBlockDefaultEncodingContext.compressAndEncrypt(HFileBlockDefaultEncodingContext.java:132)
    at org.apache.hadoop.hbase.io.hfile.HFileBlock$Writer.finishBlock(HFileBlock.java:989)
    at org.apache.hadoop.hbase.io.hfile.HFileBlock$Writer.ensureBlockReady(HFileBlock.java:961)
    at org.apache.hadoop.hbase.io.hfile.HFileBlock$Writer.finishBlockAndWriteHeaderAndData(HFileBlock.java:1077)
```

6. Perform the same operations on other threads. It is found that the threads are compaction threads.



```
"regionserver/ahbd-hbase-dat1/12.2.168.1:21302-longCompactions-1482676601473" #1629 prio=5 os_prio=0 tid=0x00007fa61454d800 nid=0x77a0 runnable
   java.lang.Thread.State: RUNNABLE
    at org.apache.hadoop.hdfs.DFSOutputStream.writeChunk(DFSOutputStream.java:425)
    - locked <0x000000020276ba38> (a org.apache.hadoop.hdfs.DFSOutputStream)
    at org.apache.hadoop.fs.FSOutputSummer.writeChecksumChunks(FSOutputSummer.java:214)
    at org.apache.hadoop.fs.FSOutputSummer.flushBuffer(FSOutputSummer.java:165)
    - locked <0x000000020276ba38> (a org.apache.hadoop.hdfs.DFSOutputStream)
    at org.apache.hadoop.fs.FSOutputSummer.flushBuffer(FSOutputSummer.java:146)
    - eliminated <0x000000020276ba38> (a org.apache.hadoop.hdfs.DFSOutputStream)
    at org.apache.hadoop.fs.FSOutputSummer.write1(FSOutputSummer.java:137)
    at org.apache.hadoop.fs.FSOutputSummer.write(FSOutputSummer.java:112)
    - locked <0x000000020276ba38> (a org.apache.hadoop.hdfs.DFSOutputStream)
    at org.apache.hadoop.fs.FSDataOutputStream$PositionCache.write(FSDataOutputStream.java:58)
    at java.io.DataOutputStream.write(DataOutputStream.java:107)
    - locked <0x00000004de9535c8> (a org.apache.hadoop.hdfs.client.HdfsDataOutputStream)
    at java.io.FilterOutputStream.write(FilterOutputStream.java:97)
```

Solution

This is a normal phenomenon.

The threads that consume a large number of CPU resources are compaction threads. Some threads invoke the Snappy compression algorithm, and some threads invoke HDFS data writing and reading. Each region has massive sets of data and numerous data files and uses the Snappy compression algorithm. For this reason, the compaction operations consume a large number of CPU resources.

Fault Locating Methods

Step 1 Run the **top** command to check the process with high CPU usage.

Step 2 Check the threads with high CPU usage in the process.

Run the **top -H -p** *<PID>* command to print CPU usage of threads under the process.

Obtain the thread with the highest CPU usage from the query result. You can also obtain the thread by running the following command:

Or run the **ps -mp** *<PID>* **-o THREAD,tid,time | sort -rn** command.

View the command output to obtain the ID of the thread with the highest CPU usage.

Step 3 Obtain the stack of the faulty thread.

The jstack tool is the most effective and reliable tool for locating Java problems.

You can obtain the jstack tool from the **java/bin** directory.

jstack <PID> > allstack.txt

Obtain the process stack and output it to a local file.

Step 4 Convert the thread ID into the hexadecimal format:

printf "%x\n" <PID>

The process ID in the command output is the TID.

Step 5 Run the following command to obtain the TID and output it to a local file:

jstack <PID> | grep <TID> > Onestack.txt

If you want to view the TID in the CLI only, run the following command:

jstack <PID> | grep <TID> -A 30

-A 30 indicates that 30 lines are displayed.

----End

15.8.11 HBase Failed to Started with "FileNotFoundException" in RegionServer Logs

Symptom

HBase fails to start, and the RegionServer stays in the **Restoring** state.

Cause Analysis

1. Check the RegionServer log (**/var/log/Bigdata/hbase/rs/hbase-omm-*XXX*.out**). It is found that the following information is printed:

```
 | ERROR | RS_OPEN_REGION-ab-dn01:21302-2 | ABORTING region server ab-
dn01,21302,1487663269375: The coprocessor
org.apache.kylin.storage.hbase.cube.v2.coprocessor.endpoint.CubeVisitService threw
java.io.FileNotFoundException: File does not exist: hdfs://hacluster/kylin/kylin_metadata/coprocessor/
kylin-coprocessor-1.6.0-SNAPSHOT-0.jar |
org.apache.hadoop.hbase.regionserver.HRegionServer.abort(HRegionServer.java:2123)
java.io.FileNotFoundException: File does not exist: hdfs://hacluster/kylin/kylin_metadata/coprocessor/
kylin-coprocessor-1.6.0-SNAPSHOT-0.jar
at org.apache.hadoop.hdfs.DistributedFileSystem$25.doCall(DistributedFileSystem.java:1399)
at org.apache.hadoop.hdfs.DistributedFileSystem$25.doCall(DistributedFileSystem.java:1391)
at org.apache.hadoop.fs.FileSystemLinkResolver.resolve(FileSystemLinkResolver.java:81)
at org.apache.hadoop.hdfs.DistributedFileSystem.getFileStatus(DistributedFileSystem.java:1391)
at org.apache.hadoop.fs.FileUtil.copy(FileUtil.java:340)
at org.apache.hadoop.fs.FileUtil.copy(FileUtil.java:292)
at org.apache.hadoop.fs.FileSystem.copyToLocalFile(FileSystem.java:2038)
at org.apache.hadoop.fs.FileSystem.copyToLocalFile(FileSystem.java:2007)
at org.apache.hadoop.fs.FileSystem.copyToLocalFile(FileSystem.java:1983)
at org.apache.hadoop.hbase.util.CoprocessorClassLoader.init(CoprocessorClassLoader.java:168)
at org.apache.hadoop.hbase.util.CoprocessorClassLoader.getClassLoader(CoprocessorClassLoader.java:
250)
at org.apache.hadoop.hbase.coprocessor.CoprocessorHost.load(CoprocessorHost.java:224)
at
org.apache.hadoop.hbase.regionserver.RegionCoprocessorHost.loadTableCoprocessors(RegionCoprocess
orHost.java:365)
at org.apache.hadoop.hbase.regionserver.RegionCoprocessorHost.<init>(RegionCoprocessorHost.java:
227)
at org.apache.hadoop.hbase.regionserver.HRegion.<init>(HRegion.java:783)
at org.apache.hadoop.hbase.regionserver.HRegion.<init>(HRegion.java:689)
at sun.reflect.GeneratedConstructorAccessor22.newInstance(Unknown Source)
at
sun.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:
45)
at java.lang.reflect.Constructor.newInstance(Constructor.java:423)
at org.apache.hadoop.hbase.regionserver.HRegion.newHRegion(HRegion.java:6312)
at org.apache.hadoop.hbase.regionserver.HRegion.openHRegion(HRegion.java:6622)
at org.apache.hadoop.hbase.regionserver.HRegion.openHRegion(HRegion.java:6594)
at org.apache.hadoop.hbase.regionserver.HRegion.openHRegion(HRegion.java:6550)
at org.apache.hadoop.hbase.regionserver.HRegion.openHRegion(HRegion.java:6501)
at
org.apache.hadoop.hbase.regionserver.handler.OpenRegionHandler.openRegion(OpenRegionHandler.ja
va:363)
at org.apache.hadoop.hbase.regionserver.handler.OpenRegionHandler.process(OpenRegionHandler.java:
129)
at org.apache.hadoop.hbase.executor.EventHandler.run(EventHandler.java:129)
at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1142)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:617)
at java.lang.Thread.run(Thread.java:745)
```

2. Run the **hdfs** command on the client. It is found that the following file does not exist:

   hdfs://hacluster/kylin/kylin_metadata/coprocessor/kylin-coprocessor-1.6.0-SNAPSHOT-0.jar

3. When configuring the coprocessor for HBase, make sure that the path of the corresponding JAR package is correct. Otherwise, HBase cannot be started.

Solution

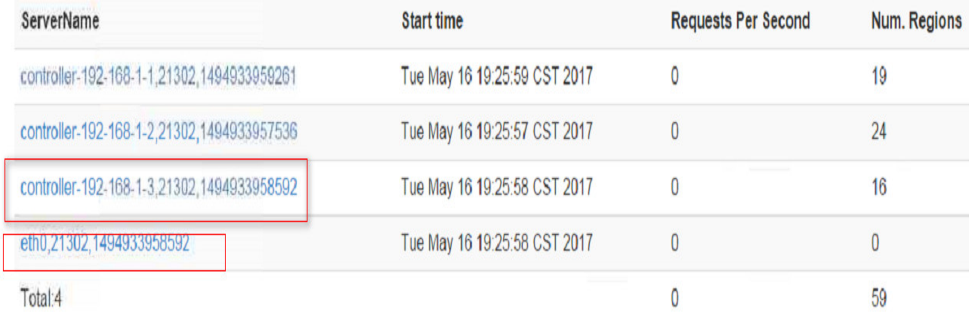
Use the Apache Kylin engine to interconnect with MRS and make sure that the JAR file of the Kylin engine exists.

15.8.12 The Number of RegionServers Displayed on the Native Page Is Greater Than the Actual Number After HBase Is Started

Symptom

After HBase is started, the number of RegionServers displayed on the HMaster native page is greater than the actual number.

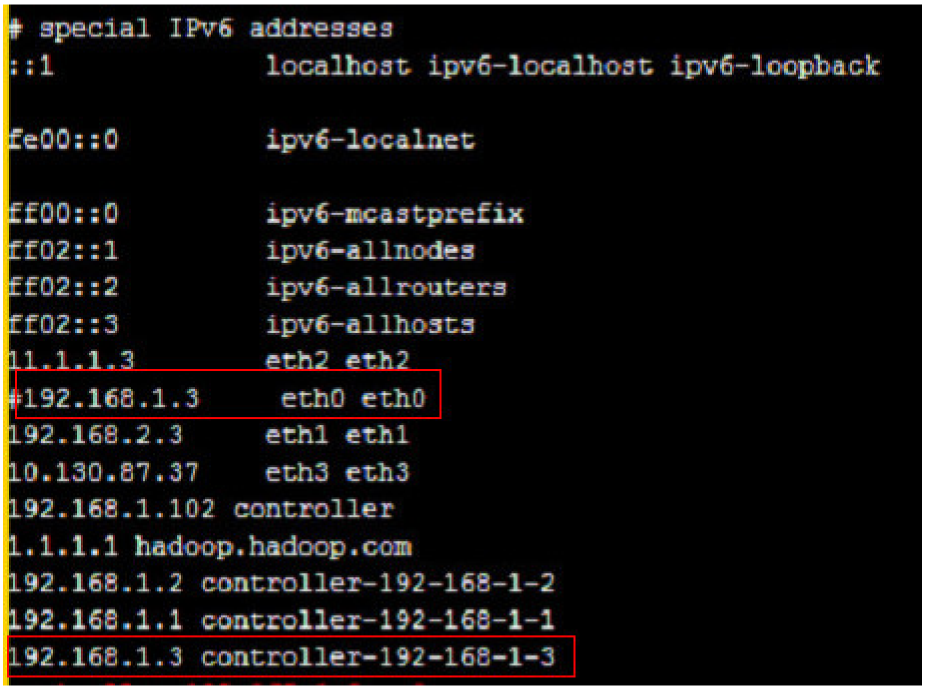
The HMaster native page shows that four RegionServers are online, as shown in the following figure.



ServerName	Start time	Requests Per Second	Num. Regions
controller-192-168-1-1,21302,1494933959261	Tue May 16 19:25:59 CST 2017	0	19
controller-192-168-1-2,21302,1494933957536	Tue May 16 19:25:57 CST 2017	0	24
controller-192-168-1-3,21302,1494933958592	Tue May 16 19:25:58 CST 2017	0	16
eth0,21302,1494933958592	Tue May 16 19:25:58 CST 2017	0	0
Total:4		0	59

Cause Analysis

As shown in the following figure, the hostname of the node in the third row is **controller-192-168-1-3** and that of the fourth row is **eth0**. The two carry the same information reported by RegionServer. Then, log in to the corresponding nodes to check the **/etc/hosts** file. It is found that the same IP address is configured for the two hostnames. For details, see the following figure:



```
# special IPv6 addresses
::1             localhost ipv6-localhost ipv6-loopback

fe00::0         ipv6-localnet

ff00::0         ipv6-mcastprefix
ff02::1         ipv6-allnodes
ff02::2         ipv6-allrouters
ff02::3         ipv6-allhosts
11.1.1.3        eth2 eth2
#192.168.1.3     eth0 eth0
192.168.2.3     eth1 eth1
10.130.87.37    eth3 eth3
192.168.1.102 controller
1.1.1.1 hadoop.hadoop.com
192.168.1.2 controller-192-168-1-2
192.168.1.1 controller-192-168-1-1
192.168.1.3 controller-192-168-1-3
```

Solution

Log in to the node where RegionServer resides, and modify the **/etc/hosts** file. Make sure that the same IP address can correspond to only one hostname.

15.8.13 RegionServer Instance Is in the Restoring State

Symptom

HBase fails to start, and the RegionServer stays in the **Restoring** state.

Cause Analysis

Check the running log (**/var/log/Bigdata/hbase/rs/hbase-omm-*XXX*.log**) of the abnormal RegionServer instance. It is found that the following information is displayed: **ClockOutOfSyncException..., Reported time is too far out of sync with master**.

```
2017-09-18 11:16:23,636 | FATAL | regionserver21302 | Master rejected startup because clock is out of sync | 
org.apache.hadoop.hbase.regionserver.HRegionServer.reportForDuty(HRegionServer.java:2059)
org.apache.hadoop.hbase.ClockOutOfSyncException: org.apache.hadoop.hbase.ClockOutOfSyncException: 
Server nl-bi-fi-datanode-24-65,21302,1505726180086 has been rejected; Reported time is too far out of 
sync with master.  Time difference of 152109ms > max allowed of 30000ms
at org.apache.hadoop.hbase.master.ServerManager.checkClockSkew(ServerManager.java:354)
...

...
2017-09-18 11:16:23,858 | ERROR | main | Region server exiting | 
org.apache.hadoop.hbase.regionserver.HRegionServerCommandLine.start(HRegionServerCommandLine.java:
70)
java.lang.RuntimeException: HRegionServer Aborted
```

This log indicates that the time difference between the abnormal RegionServer instance and the HMaster instance is greater than the allowed time difference 30s (specified by the **hbase.regionserver.maxclockskew** parameter and the default value is **30000 ms**). As a result, the RegionServer instance is abnormal.

Solution

Adjust the node time to ensure that the time difference between nodes is less than 30s.

15.8.14 HBase Failed to Start in a Newly Installed Cluster

Symptom

HBase of a newly installed cluster fails to start. The RegionServer log contains the following error information:

```
2018-02-24 16:53:03,863 | ERROR | regionserver/host3/187.6.71.69:21302 | Master passed us a different
hostname to use; was=host3, but now=187-6-71-69 |
org.apache.hadoop.hbase.regionserver.HRegionServer.handleReportForDutyResponse(HRegionServer.java:
1386)
```

Cause Analysis

In the **/etc/hosts** file, an IP address maps multiple hostnames.

Solution

Step 1 Modify the mapping between the IP address and hostnames in the **/etc/host** file.

Step 2 Restart HBase.

----End

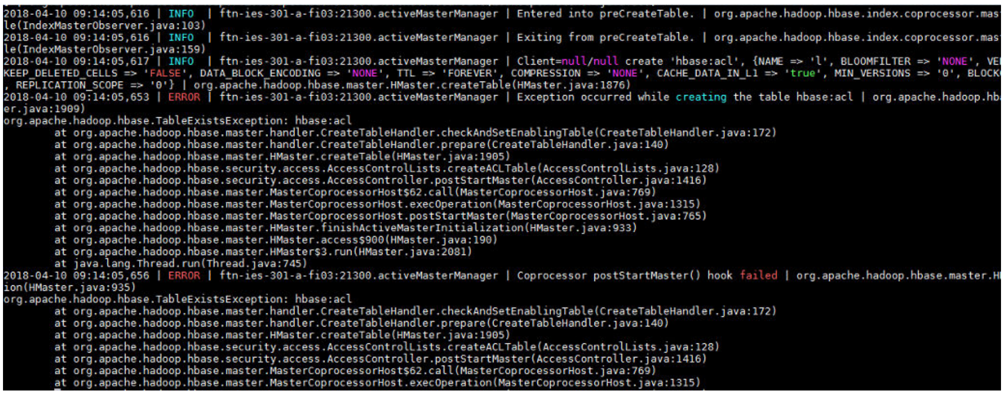
15.8.15 HBase Failed to Start Due to the Loss of the ACL Table Directory

Symptom

The HBase cluster fails to start.

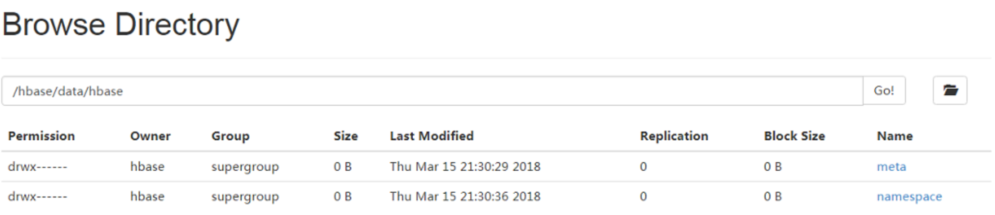
Cause Analysis

1. Check the HMaster log of HBase. The following error information is displayed:



```
2018-04-10 09:14:05,616 | INFO  | ftn-ies-301-a-fi03:21300.activeMasterManager | Entered into preCreateTable. | org.apache.hadoop.hbase.index.coprocessor.mas
le(IndexMasterObserver.java:103)
2018-04-10 09:14:05,616 | INFO  | ftn-ies-301-a-fi03:21300.activeMasterManager | Exiting from preCreateTable. | org.apache.hadoop.hbase.index.coprocessor.mas
le(IndexMasterObserver.java:159)
2018-04-10 09:14:05,617 | INFO  | ftn-ies-301-a-fi03:21300.activeMasterManager | Client=null/null create 'hbase:acl', {NAME => 'l', BLOOMFILTER => 'NONE', VE
KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', CACHE_DATA_IN_L1 => 'true', MIN_VERSIONS => '0', BLOCK
, REPLICATION_SCOPE => '0'} | org.apache.hadoop.hbase.master.HMaster.createTable(HMaster.java:1876)
2018-04-10 09:14:05,653 | ERROR | ftn-ies-301-a-fi03:21300.activeMasterManager | Exception occurred while creating the table hbase:acl | org.apache.hadoop.hb
er.java:1909)
org.apache.hadoop.hbase.TableExistsException: hbase:acl
        at org.apache.hadoop.hbase.master.handler.CreateTableHandler.checkAndSetEnablingTable(CreateTableHandler.java:172)
        at org.apache.hadoop.hbase.master.handler.CreateTableHandler.prepare(CreateTableHandler.java:140)
        at org.apache.hadoop.hbase.master.HMaster.createTable(HMaster.java:1905)
        at org.apache.hadoop.hbase.security.access.AccessControlLists.createACLTable(AccessControlLists.java:128)
        at org.apache.hadoop.hbase.security.access.AccessController.postStartMaster(AccessController.java:1416)
        at org.apache.hadoop.hbase.master.MasterCoprocessorHost$62.call(MasterCoprocessorHost.java:769)
        at org.apache.hadoop.hbase.master.MasterCoprocessorHost.execOperation(MasterCoprocessorHost.java:1315)
        at org.apache.hadoop.hbase.master.MasterCoprocessorHost.postStartMaster(MasterCoprocessorHost.java:765)
        at org.apache.hadoop.hbase.master.HMaster.finishActiveMasterInitialization(HMaster.java:933)
        at org.apache.hadoop.hbase.master.HMaster.access$900(HMaster.java:190)
        at org.apache.hadoop.hbase.master.HMaster$3.run(HMaster.java:2081)
        at java.lang.Thread.run(Thread.java:745)
2018-04-10 09:14:05,656 | ERROR | ftn-ies-301-a-fi03:21300.activeMasterManager | Coprocessor postStartMaster() hook failed | org.apache.hadoop.hbase.master.H
ion(HMaster.java:935)
org.apache.hadoop.hbase.TableExistsException: hbase:acl
        at org.apache.hadoop.hbase.master.handler.CreateTableHandler.checkAndSetEnablingTable(CreateTableHandler.java:172)
        at org.apache.hadoop.hbase.master.handler.CreateTableHandler.prepare(CreateTableHandler.java:140)
        at org.apache.hadoop.hbase.master.HMaster.createTable(HMaster.java:1905)
        at org.apache.hadoop.hbase.security.access.AccessControlLists.createACLTable(AccessControlLists.java:128)
        at org.apache.hadoop.hbase.security.access.AccessController.postStartMaster(AccessController.java:1416)
        at org.apache.hadoop.hbase.master.MasterCoprocessorHost$62.call(MasterCoprocessorHost.java:769)
        at org.apache.hadoop.hbase.master.MasterCoprocessorHost.execOperation(MasterCoprocessorHost.java:1315)
```

2. The HBase directory in HDFS is checked, which shows that the ACL table directory is lost.



Browse Directory

/hbase/data/hbase

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwx------	hbase	supergroup	0 B	Thu Mar 15 21:30:29 2018	0	0 B	meta
drwx------	hbase	supergroup	0 B	Thu Mar 15 21:30:36 2018	0	0 B	namespace

Solution

Step 1 Stop HBase.

Step 2 Log in to the HBase client as the **hbase** user and run the following command.

Example:

```
hadoop03:~ # source /opt/client/bigdata_env
hadoop03:~ # kinit hbase
Password for hbase@HADOOP.COM:
hadoop03:~ # hbase zkcli
```

Step 3 Delete the ACL table information from the ZooKeeper.

Example:

```
[zk: hadoop01:24002,hadoop02:24002,hadoop03:24002(CONNECTED) 0] deleteall /hbase/table/hbase:acl
[zk: hadoop01:24002,hadoop02:24002,hadoop03:24002(CONNECTED) 0] deleteall /hbase/table-lock/
hbase:acl
```

Step 4 Start HBase.

----End

15.8.16 HBase Failed to Start After the Cluster Is Powered Off and On

Symptom

After the ECS in the cluster is stopped and restarted, HBase fails to start.

Cause Analysis

Check the HMaster run logs. A large number of errors are reported, as shown below:

```
2018-03-26 11:10:54,185 | INFO  | hadoopc1h3,21300,1522031630949_splitLogManager__ChoreService_1 |
total tasks = 1 unassigned = 0 tasks={/hbase/splitWAL/WALs%2Fhadoopc1h1%2C213
02%2C1520214023667-splitting%2Fhadoopc1h1%252C21302%252C1520214023667.default.
1520584926990=last_update = 1522033841041 last_version = 34255 cur_worker_name =
hadoopc1h3,21302,
1520943011826 status = in_progress incarnation = 3 resubmits = 3 batch = installed = 1 done = 0 error = 0}
| org.apache.hadoop.hbase.master.SplitLogManager$TimeoutMonitor.chore
(SplitLogManager.java:745)
2018-03-26 11:11:00,185 | INFO  | hadoopc1h3,21300,1522031630949_splitLogManager__ChoreService_1 |
total tasks = 1 unassigned = 0 tasks={/hbase/splitWAL/WALs%2Fhadoopc1h1%2C213
02%2C1520214023667-splitting%2Fhadoopc1h1%252C21302%252C1520214023667.default.
1520584926990=last_update = 1522033841041 last_version = 34255 cur_worker_name =
hadoopc1h3,21302,
1520943011826 status = in_progress incarnation = 3 resubmits = 3 batch = installed = 1 done = 0 error = 0}
| org.apache.hadoop.hbase.master.SplitLogManager$TimeoutMonitor.chore
(SplitLogManager.java:745)
```

```
2018-03-26 11:11:06,185 | INFO  | hadoopc1h3,21300,1522031630949_splitLogManager__ChoreService_1 |
total tasks = 1 unassigned = 0 tasks={/hbase/splitWAL/WALs%2Fhadoopc1h1%2C213
02%2C1520214023667-splitting%2Fhadoopc1h1%252C21302%252C1520214023667.default.
1520584926990=last_update = 1522033841041 last_version = 34255 cur_worker_name =
hadoopc1h3,21302,
1520943011826 status = in_progress incarnation = 3 resubmits = 3 batch = installed = 1 done = 0 error = 0}
| org.apache.hadoop.hbase.master.SplitLogManager$TimeoutMonitor.chore
(SplitLogManager.java:745)
2018-03-26 11:11:10,787 | INFO  | RpcServer.reader=9,bindAddress=hadoopc1h3,port=21300 | Kerberos
principal name is hbase/hadoop.hadoop.com@HADOOP.COM | org.apache.hadoop.hbase
.ipc.RpcServer$Connection.readPreamble(RpcServer.java:1532)
2018-03-26 11:11:12,185 | INFO  | hadoopc1h3,21300,1522031630949_splitLogManager__ChoreService_1 |
total tasks = 1 unassigned = 0 tasks={/hbase/splitWAL/WALs%2Fhadoopc1h1%2C213
02%2C1520214023667-splitting%2Fhadoopc1h1%252C21302%252C1520214023667.default.
1520584926990=last_update = 1522033841041 last_version = 34255 cur_worker_name =
hadoopc1h3,21302,
1520943011826 status = in_progress incarnation = 3 resubmits = 3 batch = installed = 1 done = 0 error = 0}
| org.apache.hadoop.hbase.master.SplitLogManager$TimeoutMonitor.chore
(SplitLogManager.java:745)
2018-03-26 11:11:18,185 | INFO  | hadoopc1h3,21300,1522031630949_splitLogManager__ChoreService_1 |
total tasks = 1 unassigned = 0 tasks={/hbase/splitWAL/WALs%2Fhadoopc1h1%2C213
02%2C1520214023667-splitting%2Fhadoopc1h1%252C21302%252C1520214023667.default.
1520584926990=last_update = 1522033841041 last_version = 34255 cur_worker_name =
hadoopc1h3,21302,
1520943011826 status = in_progress incarnation = 3 resubmits = 3 batch = installed = 1 done = 0 error = 0}
| org.apache.hadoop.hbase.master.SplitLogManager$TimeoutMonitor.chore
(SplitLogManager.java:745)
```

The WAL splitting of RegionServer fails when the node is powered on and off.

Solution

Step 1 Stop HBase.

Step 2 Run the **hdfs fsck** command to check the health status of the **/hbase/WALs** file.

hdfs fsck /hbase/WALs

If the following command output is displayed, all files are normal. If any file is abnormal, rectify the fault, and then perform the subsequent operations.

```
The filesystem under path '/hbase/WALs' is HEALTHY
```

Step 3 Back up the **/hbase/WALs** file.

hdfs dfs -mv /hbase/WALs /hbase/WALs_old

Step 4 Run the following command to create the **/hbase/WALs** directory.

hdfs dfs -mkdir /hbase/WALs

Make sure that the permission on the directory is **hbase:hadoop**.

Step 5 Start HBase.

----End

15.8.17 Failed to Import HBase Data Due to Oversized File Blocks

Symptom

Error Message "NotServingRegionException" is displayed when data is imported to HBase.

Cause Analysis

When a block is greater than 2 GB, a read exception occurs during the seek operation of the HDFS. A full GC occurs when data is frequently written to the RegionServer. As a result, the heartbeat between the HMaster and RegionServer becomes abnormal, and the HMaster marks the RegionServer as dead, and the RegionServer is forcibly restarted. After the restart, the servercrash mechanism is triggered to roll back WALs. Currently, the **splitwal** file has reached 2.1 GB and has only one block. As a result, the HDFS seek operation becomes abnormal and the WAL file splitting fails. However, the RegionServer detects that the WAL needs to be split and triggers the splitwal mechanism, causing a loop between WAL splitting and the splitting failure. In this case, the regions on the RegionServer node cannot be brought online, and an exception is thrown indicating that the region is not online when a region on the RegionServer is queried.

Procedure

Step 1 Go to the HBase service page.

- For versions earlier than MRS 2.0.1: Log in to MRS Manager and choose **Services** > **HBase**.
- For MRS 2.0.1 or later: Click the cluster name on the MRS console and choose **Components** > **HBase**.

  NOTE

  If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)
- For MRS 3.*x* or later: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the desired cluster, and choose **Services** > **HBase**.

Step 2 On the right of **HMaster Web UI**, click **HMaster (Active)** to go to the HBase Web UI page.

Step 3 On the **Procedures** page, view the node where the problem occurs.

Step 4 Log in to the faulty node as user **root** and run the **hdfs dfs -ls** command to view all block information.

Step 5 Run the **hdfs dfs -mkdir** command to create a directory for storing faulty blocks.

Step 6 Run the **hdfs dfs -mv** command to move the faulty block to the new directory.

----End

Summary and Suggestions

The following is provided for your reference:

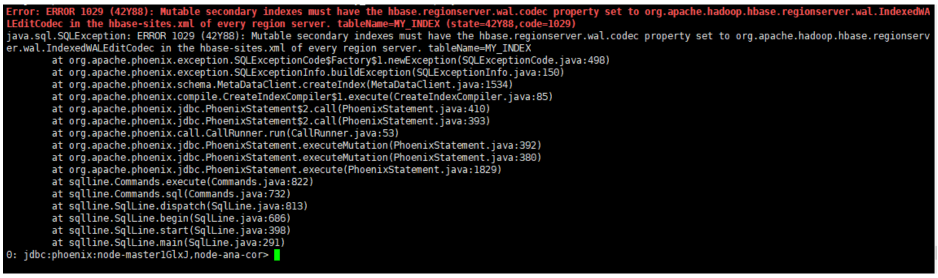
- If data blocks are corrupted, run the **hdfs fsck /tmp -files -blocks -racks** command to check the health information about data blocks.
- If you perform data operations when a region is being split, **NotServingRegionException** is thrown.

15.8.18 Failed to Load Data to the Index Table After an HBase Table Is Created Using Phoenix

Symptom

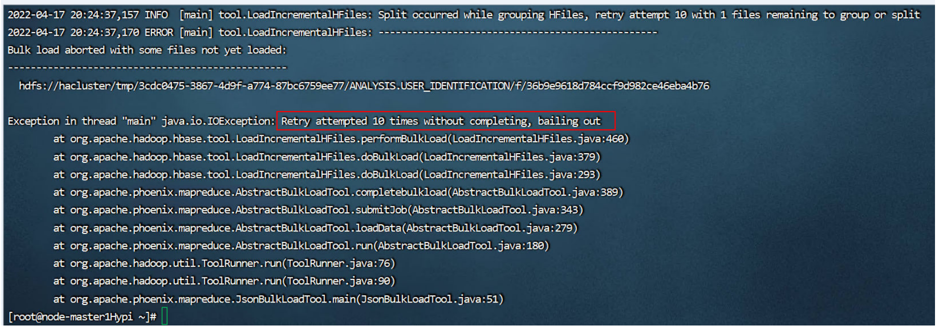
A user fails to run commands to load data to the index table after creating an HBase table using Phoenix. The following error information is displayed:

- MRS 2.x or earlier: Mutable secondary indexes must have the hbase.regionserver.wal.codec property set to org.apache.hadoop.hbase.regionserver.wal.IndexedWALEditCodec in the hbase-sites.xml of every region server. tableName=MY_INDEX (state=42Y88,code=1029)



```
Error: ERROR 1029 (42Y88): Mutable secondary indexes must have the hbase.regionserver.wal.codec property set to org.apache.hadoop.hbase.regionserver.wal.IndexedWA
LEditCodec in the hbase-sites.xml of every region server. tableName=MY_INDEX (state=42Y88,code=1029)
java.sql.SQLException: ERROR 1029 (42Y88): Mutable secondary indexes must have the hbase.regionserver.wal.codec property set to org.apache.hadoop.hbase.regionserv
er.wal.IndexedWALEditCodec in the hbase-sites.xml of every region server. tableName=MY_INDEX
        at org.apache.phoenix.exception.SQLExceptionCode$Factory$1.newException(SQLExceptionCode.java:498)
        at org.apache.phoenix.exception.SQLExceptionInfo.buildException(SQLExceptionInfo.java:150)
        at org.apache.phoenix.schema.MetaDataClient.createIndex(MetaDataClient.java:1534)
        at org.apache.phoenix.compile.CreateIndexCompiler$1.execute(CreateIndexCompiler.java:85)
        at org.apache.phoenix.jdbc.PhoenixStatement$2.call(PhoenixStatement.java:410)
        at org.apache.phoenix.jdbc.PhoenixStatement$2.call(PhoenixStatement.java:393)
        at org.apache.phoenix.call.CallRunner.run(CallRunner.java:53)
        at org.apache.phoenix.jdbc.PhoenixStatement.executeMutation(PhoenixStatement.java:392)
        at org.apache.phoenix.jdbc.PhoenixStatement.executeMutation(PhoenixStatement.java:380)
        at org.apache.phoenix.jdbc.PhoenixStatement.execute(PhoenixStatement.java:1829)
        at sqlline.Commands.execute(Commands.java:822)
        at sqlline.Commands.sql(Commands.java:732)
        at sqlline.SqlLine.dispatch(SqlLine.java:813)
        at sqlline.SqlLine.begin(SqlLine.java:686)
        at sqlline.SqlLine.start(SqlLine.java:398)
        at sqlline.SqlLine.main(SqlLine.java:291)
0: jdbc:phoenix:node-master1GlxJ,node-ana-cor>
```

- MRS 3. x or later: Exception in thread "main" java.io.IOException: Retry attempted 10 times without completing, bailing out



```
2022-04-17 20:24:37,157 INFO  [main] tool.LoadIncrementalHFiles: Split occurred while grouping HFiles, retry attempt 10 with 1 files remaining to group or split
2022-04-17 20:24:37,170 ERROR [main] tool.LoadIncrementalHFiles: -----------------------------------------------
Bulk load aborted with some files not yet loaded:
-----------------------------------------------
  hdfs://hacluster/tmp/3cdc0475-3867-4d9f-a774-87bc6759ee77/ANALYSIS.USER_IDENTIFICATION/f/36b9e9618d784ccf9d982ce46eba4b76

Exception in thread "main" java.io.IOException: Retry attempted 10 times without completing, bailing out
        at org.apache.hadoop.hbase.tool.LoadIncrementalHFiles.performBulkLoad(LoadIncrementalHFiles.java:460)
        at org.apache.hadoop.hbase.tool.LoadIncrementalHFiles.doBulkLoad(LoadIncrementalHFiles.java:379)
        at org.apache.hadoop.hbase.tool.LoadIncrementalHFiles.doBulkLoad(LoadIncrementalHFiles.java:293)
        at org.apache.phoenix.mapreduce.AbstractBulkLoadTool.completebulkload(AbstractBulkLoadTool.java:389)
        at org.apache.phoenix.mapreduce.AbstractBulkLoadTool.submitJob(AbstractBulkLoadTool.java:343)
        at org.apache.phoenix.mapreduce.AbstractBulkLoadTool.loadData(AbstractBulkLoadTool.java:279)
        at org.apache.phoenix.mapreduce.AbstractBulkLoadTool.run(AbstractBulkLoadTool.java:180)
        at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)
        at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:90)
        at org.apache.phoenix.mapreduce.JsonBulkLoadTool.main(JsonBulkLoadTool.java:51)
[root@node-master1Hypi ~]#
```

Procedure

Step 1 For MRS 2.x or earlier, perform the following operations:

1. Log in to MRS Manager as user **admin**, choose **Services**, and click **HBase**. On the **Service Configuration** tab, select **All** from the **Type** drop-down list, choose **HMaster** > **Customization**, and add a configuration item for parameter **hbase.hmaster.config.expandor** with name **hbase.regionserver.wal.codec** and value **org.apache.hadoop.hbase.regionserver.wal.IndexedWALEditCodec**.

2. Choose **RegionServer** > **Customization**, add a configuration item for parameter **hbase.regionserver.config.expandor** with name **hbase.regionserver.wal.codec** and value **org.apache.hadoop.hbase.regionserver.wal.IndexedWALEditCodec**, and click **Save Configuration**. Then enter the password of the current user and click **OK**.
3. On the **Service Status** page, click **More** and select **Restart Service**. Enter the password of the current user and click **OK** to restart the HBase service.

Step 2 For MRS 3.x or later, perform the following operations:

1. Log in to FusionInsight Manager as user **admin** and choose **Cluster** > **Services** > **HBase**. On the HBase page, choose **Configurations** > **All Configurations** > **RegionServer** > **Customization**. In the right pane, add a configuration item for parameter **hbase.regionserver.config.expandor** with name **hbase.regionserver.wal.codec** and value **org.apache.hadoop.hbase.regionserver.wal.IndexedWALEditCodec**.
2. Choose **HMaster** > **Customization**, and add a configuration item for parameter **hbase.hmaster.config.expandor** with name **hbase.regionserver.wal.codec** and value **org.apache.hadoop.hbase.regionserver.wal.IndexedWALEditCodec**.
3. Click **Save**. In the dialog box that is displayed, click **OK** to save the configuration.
4. On the **Dashboard** page, click **More** and select **Restart Service**. Enter the password of the current user and click **OK** to restart the HBase service.

----End

15.8.19 Failed to Run the hbase shell Command on the MRS Cluster Client

Issue

A user fails to run the **hbase shell** command on the MRS cluster client.

Cause Analysis

- Environment variables have not been configured before the **hbase shell** command is executed.
- The HBase client is not installed in the MRS cluster.

Procedure

Step 1 Log in to the node where the client is installed as user **root**, switch to the client installation directory, and check whether the HBase client is installed.

- If yes, go to **Step 2**.
- If no, download and install the client.

Step 2 Run the following command to set environment variables:

source bigdata_env

Step 3 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the current user. The current user must have the permission to create HBase tables. If Kerberos authentication is disabled for the current cluster, skip this step.

kinit *MRS cluster user*

Step 4 Run the HBase client command.

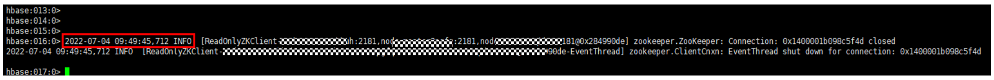
hbase shell

----End

15.8.20 Disordered Information Display on the HBase Shell Client Console Due to Printing of the INFO Information

Issue

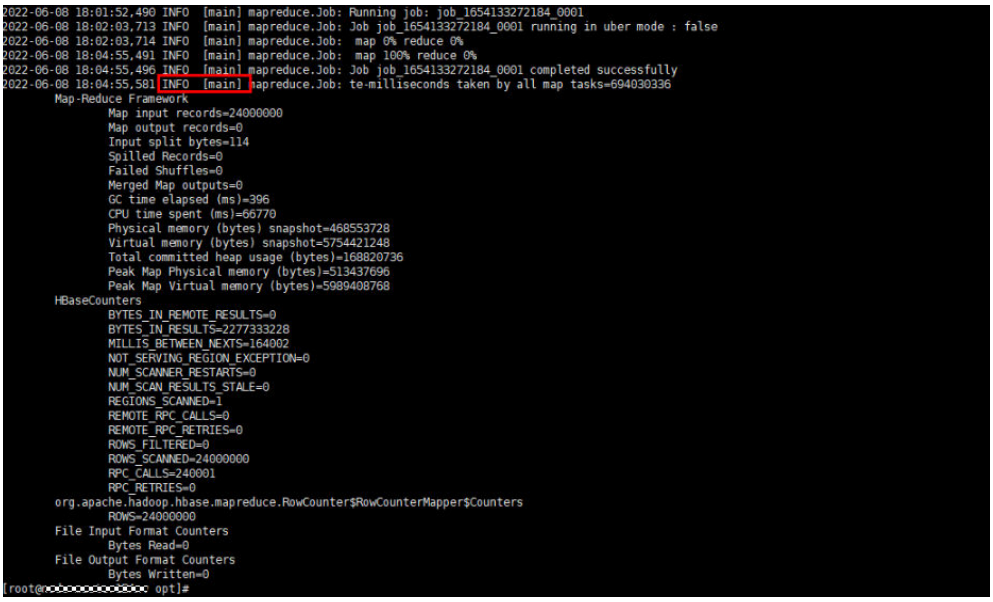
When the HBase shell is used, INFO information is printed on the console, causing disordered information display. Before entering a command, the user has to press **Enter** to clean the console. The following figure provides an example.



```
hbase:013:0>
hbase:014:0>
hbase:015:0>
hbase:016:0> 2022-07-04 09:49:45,712 INFO  [ReadOnlyZKClient-[redacted]uh:2181,nod[redacted]:2181,nod[redacted]181@0x284990de] zookeeper.ZooKeeper: Connection: 0x1400001b098c5f4d closed
2022-07-04 09:49:45,712 INFO  [ReadOnlyZKClient-[redacted]90de-EventThread] zookeeper.ClientCnxn: EventThread shut down for connection: 0x1400001b098c5f4d

hbase:017:0>
```

Cause Analysis

- By default, the log printing function of the HBase client is set to **INFO,console**. Therefore, INFO logs will be printed to the console, affecting the information display on the HBase shell console.
- The HBase client supports many commands, such as **hbase shell**, **hbase hbck**, and **hbase org.apache.hadoop.hbase.mapreduce.RowCounter**, and will support more in the future. The output type of some commands is **INFO**. If **INFO** is disabled, the output of these commands will be lost. For example, the output of **RowCounter** is of the **INFO** type.



```
2022-06-08 18:01:52,490 INFO  [main] mapreduce.Job: Running job: job_1654133272184_0001
2022-06-08 18:02:03,713 INFO  [main] mapreduce.Job: Job job_1654133272184_0001 running in uber mode : false
2022-06-08 18:02:03,714 INFO  [main] mapreduce.Job:  map 0% reduce 0%
2022-06-08 18:04:55,491 INFO  [main] mapreduce.Job:  map 100% reduce 0%
2022-06-08 18:04:55,496 INFO  [main] mapreduce.Job: Job job_1654133272184_0001 completed successfully
2022-06-08 18:04:55,581 INFO  [main] mapreduce.Job: te-milliseconds taken by all map tasks=694030336
        Map-Reduce Framework
                Map input records=24000000
                Map output records=0
                Input split bytes=114
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=396
                CPU time spent (ms)=66770
                Physical memory (bytes) snapshot=468553728
                Virtual memory (bytes) snapshot=5754421248
                Total committed heap usage (bytes)=168820736
                Peak Map Physical memory (bytes)=513437696
                Peak Map Virtual memory (bytes)=5989408768
        HBaseCounters
                BYTES_IN_REMOTE_RESULTS=0
                BYTES_IN_RESULTS=2277333228
                MILLIS_BETWEEN_NEXTS=164002
                NOT_SERVING_REGION_EXCEPTION=0
                NUM_SCANNER_RESTARTS=0
                NUM_SCAN_RESULTS_STALE=0
                REGIONS_SCANNED=1
                REMOTE_RPC_CALLS=0
                REMOTE_RPC_RETRIES=0
                ROWS_FILTERED=0
                ROWS_SCANNED=24000000
                RPC_CALLS=240001
                RPC_RETRIES=0
        org.apache.hadoop.hbase.mapreduce.RowCounter$RowCounterMapper$Counters
                ROWS=24000000
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
[root@[redacted] opt]#
```

Procedure

Step 1 Log in to the node where the HBase client is installed as user **root**.

Step 2 Add the following information to the *HBase client installation directory***/HBase/component_env** file:

export HBASE_ROOT_LOGGER=INFO,RFA

Logs are exported to log files. If you run the **hbase org.apache.hadoop.hbase.mapreduce.RowCounter** command, you can view the execution result in the *HBase client installation directory***/HBase/hbase/logs/hbase.log** file.

Step 3 Switch to the HBase client installation directory and run the following commands for the configuration to take effect:

cd *HBase client installation directory*

source HBase/component_env

----End

15.8.21 HBase Failed to Start Due to Insufficient RegionServer Memory

Issue

The HBase service fails to start because the remaining RegionServer memory is insufficient.

Cause Analysis

The troubleshooting process is as follows:

1. Log in to the master node, go to the **/var/log/Bigdata** directory, and search for the HBase log. The log contains error message "connect regionserver timeout".
2. Log in to the RegionServer node in 1 that cannot be connected to HMaster and go to the **/var/log/Bigdata** directory to search for the HBase log. The RegionServer reports error message "error='Cannot allocate memory'(errno=12)".
3. According to the error message in 2, the startup failure is caused by insufficient RegionServer memory.

Procedure

Step 1 Log in to the RegionServer node where the error is reported and run the following command to check the remaining memory of the node:

free -g

Step 2 Run the **top** command to check the memory usage of the node.

Step 3 Stop the memory-consuming processes (not the processes of the MRS components) as prompted and restart the HBase service.

NOTE

Besides MRS components, jobs on Yarn are allocated to core nodes in the cluster, thereby occupying node memory. If the startup failure is caused by memory-consuming Yarn jobs, you are advised to expand the capacity of core nodes.

----End

15.9 Using HDFS

15.9.1 All NameNodes Become the Standby State After the NameNode RPC Port of HDFS Is Changed

Issue

After the NameNode RPC port is changed on the page and HDFS is restarted, all NameNodes are in the standby state, causing a cluster exception.

Symptom

All NameNodes are in the standby state, causing a cluster exception.

Cause Analysis

After the cluster is installed and started, if the NameNode RPC port is changed, the Zkfc service must be formatted to update node information on ZooKeeper.

Procedure

Step 1 Log in to Manager and stop the HDFS service.

NOTE

Do not stop related services when stopping HDFS.

Step 2 After the services are stopped, log in to the Master node whose RPC port is changed.

NOTE

If the RPC port is changed on both Master nodes, you can log in to either of the Master nodes.

Step 3 Run the **su - omm** command to switch to user **omm**.

NOTE

For a security cluster, run the **kinit hdfs** command for authentication.

Step 4 Run the following command to load the environment variable script to the environment:

cd ${BIGDATA_HOME}/*MRS_X.X.X***/***1_8_***Zkfc/etc**

source ${BIGDATA_HOME}/*MRS_X.X.X***/install/FusionInsight-Hadoop-3.1.1/hadoop/sbin/exportENV_VARS.sh**

NOTE

In the preceding command, *MRS_X.X.X* and *1_8* vary depending on the actual version.

Step 5 After the loading is complete, run the following command to format the Zkfc:

cd ${HADOOP_HOME}/bin

./hdfs zkfc -formatZK

Step 6 After the formatting is successful, restart HDFS on Manager.

NOTE

If the RPC port of the NameNode is changed, the configuration file must be updated for all clients that have been installed.

----End

15.9.2 An Error Is Reported When the HDFS Client Is Used After the Host Is Connected Using a Public Network IP Address

Issue

When the host is connected using a public network IP address, the HDFS client cannot be used and the message "**-bash: hdfs: command not found**" is displayed when the HDFS is running.

Symptom

When the host is connected using a public network IP address, the HDFS client cannot be used and the message "**-bash: hdfs: command not found**" is displayed when the HDFS is running.

Possible Causes

The environment variables are not set before the user logs in to the Master node and runs the command.

Procedure

Step 1 Log in to any Master node as user **root**.

Step 2 Run the **source /opt/client/bigdata_env** command to configure environment variables.

Step 3 Run the **hdfs** command to use the HDFS client.

----End

15.9.3 Failed to Use Python to Remotely Connect to the Port of HDFS

Issue

Failed to use Python to remotely connect to the port of HDFS.

Symptom

Failed to use Python to remotely connect to port 50070 of HDFS.

Cause Analysis

The default port of open source HDFS is 50070 for versions earlier than 3.0.0 and is 9870 for version 3.0.0 or later. The port used by the user does not match the HDFS version.

Step 1 Log in to the active Master node in the cluster.

Step 2 Run the **su - omm** command to switch to user **omm**.

Step 3 Run the **/opt/Bigdata/om-0.0.1/sbin/queryVersion.sh** command to check the HDFS version in the cluster.

Determine the port number of the open-source component based on the version number.

Step 4 Run the **netstat -anp|grep ${port}** command to check whether the default port number of the component exists.

If it does not exist, the default port number is changed. Change the port to the default port and reconnect to HDFS.

If it exists, contact technical support.

NOTE

- **${port}**: indicates the default port number corresponding to the component version.
- If you have changed the default port number, use the new port number to connect to HDFS. You are advised not to change the default port number.

----End

15.9.4 HDFS Capacity Usage Reaches 100%, Causing Unavailable Upper-layer Services Such as HBase and Spark

Issue

The HDFS capacity usage of the cluster reaches 100%, and the HDFS service status is read-only. As a result, upper-layer services such as HBase and Spark are unavailable.

Symptom

The HDFS capacity usage is 100%, the disk capacity usage is only about 85%, and the HDFS service status is read-only. As a result, upper-layer services such as HBase and Spark are unavailable.

Cause Analysis

Currently, NodeManager and DataNode share data disks. By default, MRS reserves 15% of data disk space for non-HDFS. You can change the percentage of data disk space by setting the HDFS parameter **dfs.datanode.du.reserved.percentage**.

If the HDFS disk usage is 100%, you can set **dfs.datanode.du.reserved.percentage** to a smaller value to restore services and then expand disk capacity.

Procedure

Step 1 Log in to any Master node in the cluster.

Step 2 Run the **source /opt/client/bigdata_env** command to initialize environment variables.

NOTE

If it is a security cluster, run the **kinit -kt <keytab file> <pricipal name>** command for authentication.

Step 3 Run the **hdfs dfs –put ./startDetail.log /tmp** command to check whether HDFS fails to write files.

```
19/05/12 10:07:32 WARN hdfs.DataStreamer: DataStreamer Exception
org.apache.hadoop.ipc.RemoteException(java.io.IOException): File /tmp/startDetail.log._COPYING_ could
only be replicated to 0 nodes instead of minReplication (=1).  There are 3 datanode(s) running and no
node(s) are excluded in this operation.
```

Step 4 Run the **hdfs dfsadmin -report** command to check the used HDFS capacity. The command output shows that the HDFS capacity usage has reached 100%.

```
Configured Capacity: 5389790579100 (4.90 TB)
Present Capacity: 5067618628404 (4.61 TB)
DFS Remaining: 133350196 (127.17 MB)
DFS Used: 5067485278208 (4.61 TB)
DFS Used%: 100.00%
Under replicated blocks: 10
Blocks with corrupt replicas: 0
Missing blocks: 0
Missing blocks (with replication factor 1): 0
Pending deletion blocks: 0
```

Step 5 When the HDFS capacity usage reaches 100%, change the percentage of data disk space by setting the HDFS parameter **dfs.datanode.du.reserved.percentage**.

1. Go to the service configuration page.
   - MRS Manager: Log in to MRS Manager and choose **Services** > **HDFS** > **Configuration**.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **HDFS** > **Configurations**.
2. Select **All Configurations** and search for **dfs.datanode.du.reserved.percentage** in the search box.

3. Change the value of this parameter to **10**.

Step 6 After the modification, increase the number of disks of the Core node.

----End

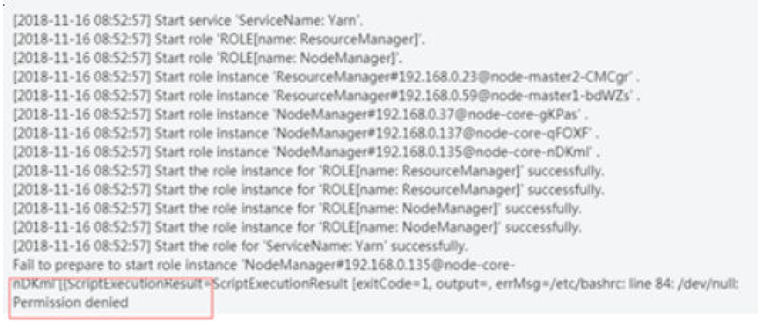
15.9.5 An Error Is Reported During HDFS and Yarn Startup

Issue

An error is reported during HDFS and Yarn startup.

Symptom

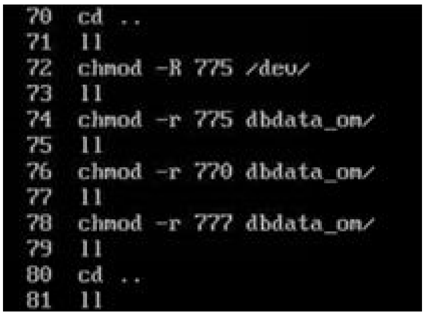
HDFS and Yarn fail to be started. The following error information is displayed: **/dev/null Permission denied**



```
[2018-11-16 08:52:57] Start service 'ServiceName: Yarn'.
[2018-11-16 08:52:57] Start role 'ROLE[name: ResourceManager]'.
[2018-11-16 08:52:57] Start role 'ROLE[name: NodeManager]'.
[2018-11-16 08:52:57] Start role instance 'ResourceManager#192.168.0.23@node-master2-CMCgr' .
[2018-11-16 08:52:57] Start role instance 'ResourceManager#192.168.0.59@node-master1-bdWZs' .
[2018-11-16 08:52:57] Start role instance 'NodeManager#192.168.0.37@node-core-gKPas' .
[2018-11-16 08:52:57] Start role instance 'NodeManager#192.168.0.137@node-core-qFOXF' .
[2018-11-16 08:52:57] Start role instance 'NodeManager#192.168.0.135@node-core-nDKml' .
[2018-11-16 08:52:57] Start the role instance for 'ROLE[name: ResourceManager]' successfully.
[2018-11-16 08:52:57] Start the role instance for 'ROLE[name: ResourceManager]' successfully.
[2018-11-16 08:52:57] Start the role instance for 'ROLE[name: NodeManager]' successfully.
[2018-11-16 08:52:57] Start the role instance for 'ROLE[name: NodeManager]' successfully.
[2018-11-16 08:52:57] Start the role for 'ServiceName: Yarn' successfully.
Fail to prepare to start role instance 'NodeManager#192.168.0.135@node-core-
nDKml' [[ScriptExecutionResult=ScriptExecutionResult [exitCode=1, output=, errMsg=/etc/bashrc: line 84: /dev/null:
Permission denied
```

Cause Analysis

The customer changed the permission value of **/dev/null** of the VM system to **775**.



```
70  cd ..
71  ll
72  chmod -R 775 /dev/
73  ll
74  chmod -r 775 dbdata_om/
75  ll
76  chmod -r 770 dbdata_om/
77  ll
78  chmod -r 777 dbdata_om/
79  ll
80  cd ..
81  ll
```

Procedure

Step 1 Log in to any Master node in the cluster as user **root**.

Step 2 After successful login, run the **chmod 666 /dev/null** command to modify the permission value of **/dev/null** to **666**.

Step 3 Run the **ls -al /dev/null** command to check whether the new permission value of **/dev/null** is **666**. If it is not, change the value to **666**.

Step 4 After the modification is successful, restart HDFS and Yarn.

----End

15.9.6 HDFS Permission Setting Error

Issue

When using MRS, a user has the permission to delete or create files in another user's HDFS directory.

Symptom

When using MRS, a user has the permission to delete or create files in another user's HDFS directory.

Cause Analysis

The user has the permission for the **ficommon** group and therefore can perform any operations on the HDFS. You need to remove the user's **ficommon** group permission.

Procedure

Step 1 Log in to the master node in the cluster as user **root**.

Step 2 Run the **id ${Username}** command to check whether the user has the **ficommon** group permission.

If the user has the **ficommon** group permission, go to **Step 3**. If the user does not have the **ficommon** group permission, contact technical support.

NOTE

${Username} indicates the name of the user whose HDFS permission is incorrectly set.

Step 3 Run the **gpasswd -d ${Username} ficommon** command to delete the user's **ficommon** group permission.

NOTE

${Username} indicates the name of the user whose HDFS permission is incorrectly set.

Step 4 Modify parameters on Manager.

MRS Manager (applicable to versions earlier than MRS 3.*x*):

1. Log in to MRS Manager and choose **Services** > **HDFS** > **Service Configuration**.
2. Set **Type** to **All**, enter **dfs.permissions.enabled** in the search box, and change the parameter value to **true**.
3. Click **Save Configuration** and restart the HDFS service.

FusionInsight Manager (applicable to MRS 3.*x* or later):

1. Log in to FusionInsight Manager. Choose **Cluster** > **Services** > **HDFS** > **Configurations** > **All Configurations**.
2. Enter **dfs.permissions.enabled** in the search box and change the value to **true**.

3. After the modification is complete, click **Save** and restart the HDFS service.

MRS console (applicable to MRS 2.0.1 or later):

1. Log in to the MRS console and choose **Components** > **HDFS** > **Service Configuration**.
2. Set **Type** to **All**, enter **dfs.permissions.enabled** in the search box, and change the parameter value to **true**.
3. Click **Save Configuration** and restart the HDFS service.

----End

15.9.7 A DataNode of HDFS Is Always in the Decommissioning State

Issue

A DataNode of HDFS is in the **Decommissioning** state for a long period of time.

Symptom

A DataNode of HDFS fails to be decommissioned (or the Core node fails to be scaled in), but the DataNode remains in the Decommissioning state.

Cause Analysis

During the decommissioning of a DataNode (or scale-in of the Core node) in HDFS, the decommissioning or scale-in task fails and the blacklist is not cleared because the Master node is restarted or the NodeAgent process exits unexpectedly. In this case, the DataNode remains in the **Decommissioning** state. The blacklist needs to be cleared manually.

Procedure

Step 1 Go to the service instance page.

MRS Manager:

Log in to MRS Manager and choose **Services** > **HDFS** > **Instance**.

FusionInsight Manager:

MRS 3.x or later: Log in to FusionInsight Manager and choose **Cluster** > **Service** > **HDFS** > **Instance**.

Log in to the MRS console and choose **Components** > **HDFS** > **Instances**.

MRS cluster 2.0.1 or later: Log in to the MRS console and choose **Components** > **HDFS** > **Instances**.

Step 2 Check the HDFS service instance status, locate the DataNode that is in the decommissioning state, and copy the IP address of the DataNode.

Step 3 Log in to the Master1 node and run the **cd ${BIGDATA_HOME}/MRS_*/1_*_NameNode/etc/** command to go to the blacklist directory.

Step 4 Run the **sed -i "/^*IP*$/d" excludeHosts** command to clear the faulty DataNode information from the blacklist. Replace the IP address in the command with the IP address of the faulty DataNode queried in **Step 2**. The IP address cannot contain spaces.

Step 5 If there are two Master nodes, perform **Step 3** and **Step 4** on Master2.

Step 6 Run the following command on the Master1 node to initialize environment variables:

source /opt/client/bigdata_env

Step 7 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If Kerberos authentication is disabled for the current cluster, skip this step:

kinit *MRS cluster user*

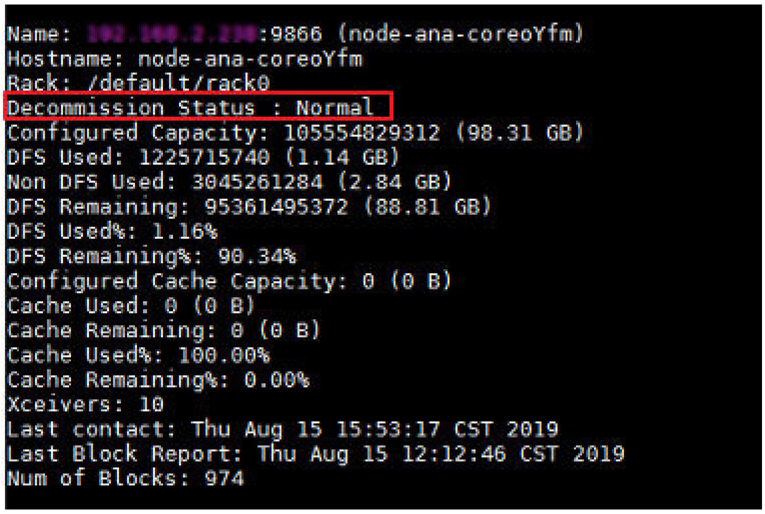
Example: **kinit admin**

Step 8 Run the following command on the Master1 node to update the HDFS blacklist:

hdfs dfsadmin -refreshNodes

Step 9 Run the **hdfs dfsadmin -report** command to check the status of each DataNode. Ensure that the DataNode corresponding to the IP address obtained in has been restored to the **Normal** state.

Figure 15-21 DataNode status



```
Name: [redacted]:9866 (node-ana-coreoYfm)
Hostname: node-ana-coreoYfm
Rack: /default/rack0
Decommission Status : Normal
Configured Capacity: 105554829312 (98.31 GB)
DFS Used: 1225715740 (1.14 GB)
Non DFS Used: 3045261284 (2.84 GB)
DFS Remaining: 95361495372 (88.81 GB)
DFS Used%: 1.16%
DFS Remaining%: 90.34%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 10
Last contact: Thu Aug 15 15:53:17 CST 2019
Last Block Report: Thu Aug 15 12:12:46 CST 2019
Num of Blocks: 974
```

Step 10 Go to the service instance page.

MRS Manager:

Log in to MRS Manager and choose **Services** > **HDFS** > **Instances**.

FusionInsight Manager:

MRS 3.x or later: Log in to FusionInsight Manager and choose **Cluster** > **Service** > **HDFS** > **Instance**.

MRS cluster 2.0.1 or later: Log in to the MRS console and choose **Components** > **HDFS** > **Instances**.

Step 11 Select the DataNode instance that is in the decommissioning state and choose **More** > **Restart Instance**.

Step 12 Wait until the restart is complete and check whether the DataNode is restored.

----End

Summary and Suggestions

Do not perform high-risk operations, such as restarting nodes, during decommissioning (or scale-in).

Related Information

None

15.9.8 HDFS Failed to Start Due to Insufficient Memory

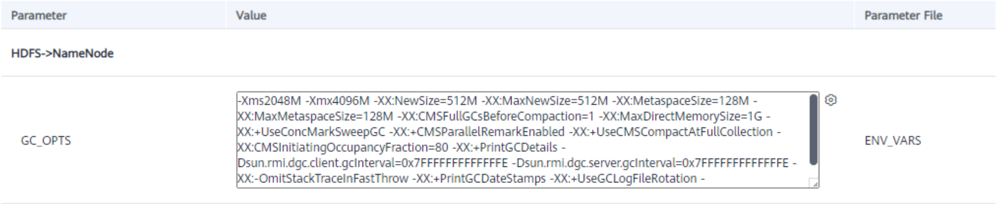
Symptom

After the HDFS service is restarted, HDFS is in the Bad state, the NameNode instance status is abnormal, and the system cannot exit the security mode for a long time.

Cause Analysis

1. In the NameNode run log (**/var/log/Bigdata/hdfs/nn/hadoop-omm-namendoe-XXX.log**), search for **WARN**. It is found that GC takes 63 seconds.
```
2017-01-22 14:52:32,641 | WARN  | org.apache.hadoop.util.JvmPauseMonitor$Monitor@1b39fd82 |
Detected pause in JVM or host machine (eg GC): pause of approximately 63750ms
GC pool 'ParNew' had collection(s): count=1 time=0ms
GC pool 'ConcurrentMarkSweep' had collection(s): count=1 time=63924ms | JvmPauseMonitor.java:189
```
2. Analyze the NameNode log **/var/log/Bigdata/hdfs/nn/hadoop-omm-namendoe-XXX.log**. It is found that the NameNode is waiting for block reporting and the total number of blocks is too large. In the following example, the total number of blocks is 36.29 million.
```
2017-01-22 14:52:32,641 | INFO  | IPC Server handler 8 on 25000 | STATE* Safe mode ON.
The reported blocks 29715437 needs additional 6542184 blocks to reach the threshold 0.9990 of total
blocks 36293915.
```
3. On Manager, check the **GC_OPTS** parameter of the NameNode:

Figure 15-22 Checking the GC_OPTS parameter of the NameNode



Parameter	Value	Parameter File
HDFS->NameNode		
GC_OPTS	-Xms2048M -Xmx4096M -XX:NewSize=512M -XX:MaxNewSize=512M -XX:MetaspaceSize=128M -XX:MaxMetaspaceSize=128M -XX:CMSFullGCsBeforeCompaction=1 -XX:MaxDirectMemorySize=1G -XX:+UseConcMarkSweepGC -XX:+CMSParallelRemarkEnabled -XX:+UseCMSCompactAtFullCollection -XX:CMSInitiatingOccupancyFraction=80 -XX:+PrintGCDetails -Dsun.rmi.dgc.client.gcInterval=0x7FFFFFFFFFFFFFE -Dsun.rmi.dgc.server.gcInterval=0x7FFFFFFFFFFFFFE -XX:-OmitStackTraceInFastThrow -XX:+PrintGCDateStamps -XX:+UseGCLogFileRotation -	ENV_VARS

4. For details about the mapping between the NameNode memory configuration and data volume, see **Table 15-2**.

Table 15-2 Mapping between NameNode memory configuration and data volume

Number of File Objects	Reference Value
10,000,000	-Xms6G -Xmx6G -XX:NewSize=512M -XX:MaxNewSize=512M
20,000,000	-Xms12G -Xmx12G -XX:NewSize=1G -XX:MaxNewSize=1G
50,000,000	-Xms32G -Xmx32G -XX:NewSize=2G -XX:MaxNewSize=3G
100,000,000	-Xms64G -Xmx64G -XX:NewSize=4G -XX:MaxNewSize=6G
200,000,000	-Xms96G -Xmx96G -XX:NewSize=8G -XX:MaxNewSize=9G
300,000,000	-Xms164G -Xmx164G -XX:NewSize=12G -XX:MaxNewSize=12G

Solution

Step 1 Modify the NameNode memory parameter based on the specifications. If the number of blocks is 36 million, change the parameter value to **-Xms32G -Xmx32G -XX:NewSize=2G -XX:MaxNewSize=3G**.

Step 2 Restart a NameNode and check that the NameNode can be started normally.

Step 3 Restart the other NameNode and check that the page status is restored.

----End

15.9.9 A Large Number of Blocks Are Lost in HDFS due to the Time Change Using ntpdate

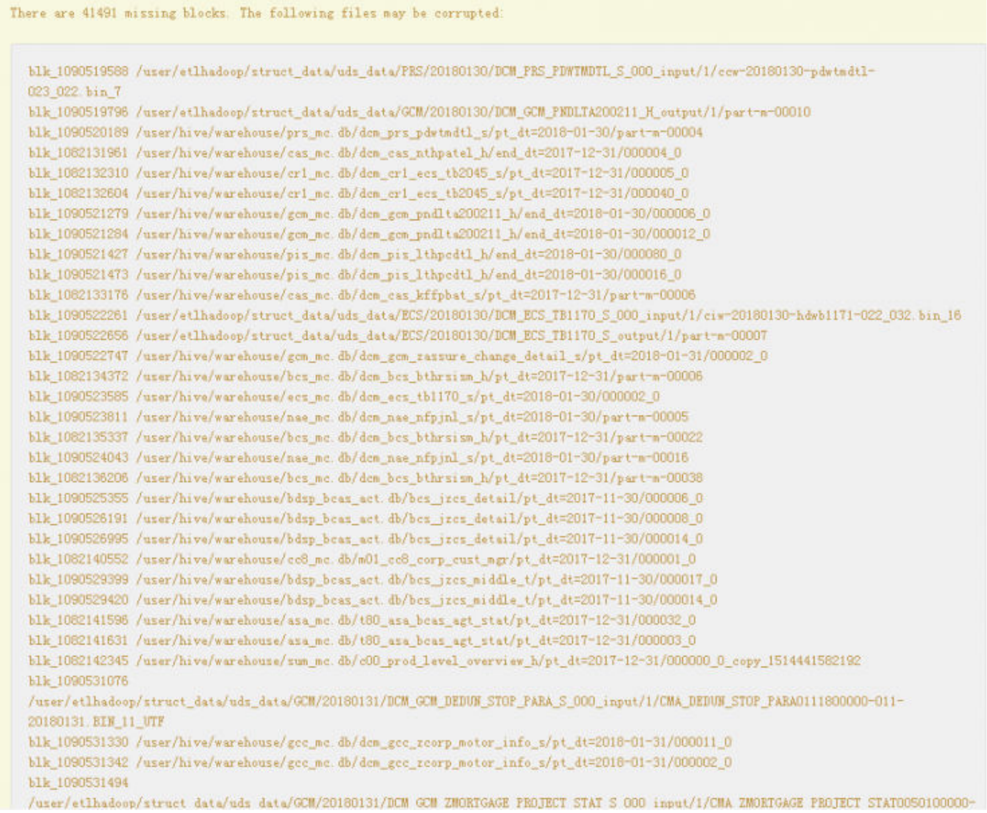
Symptom

1. A user uses **ntpdate** to change the time for a cluster that is not stopped. After the time is changed, HDFS enters the safe mode and cannot be started.
2. After the system exits the safe mode and starts, about 1 TB data is lost during the **hfck** check.

Cause Analysis

1. A large number of blocks are lost on the native NameNode page.

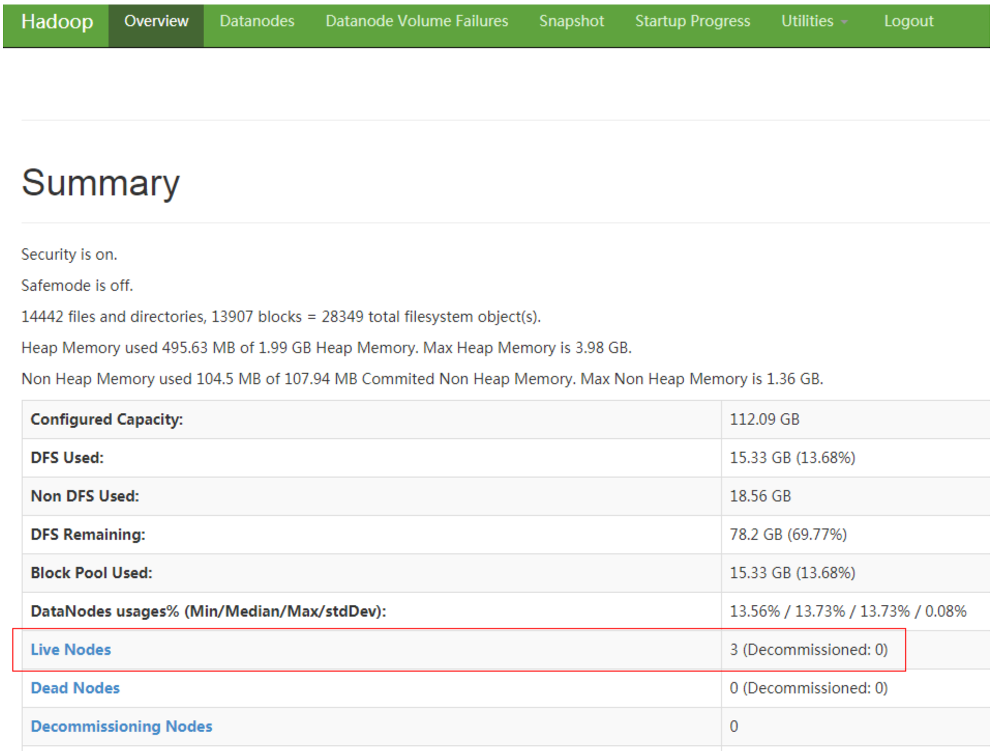
Figure 15-23 Block loss



```
There are 41491 missing blocks. The following files may be corrupted:

blk_1090519588 /user/etlhadoop/struct_data/uds_data/PRS/20180130/DCM_PRS_PDWTMDTL_S_000_input/1/ccw-20180130-pdwtmdtl-
023_022.bin_7
blk_1090519796 /user/etlhadoop/struct_data/uds_data/GCM/20180130/DCM_GCM_PNDLTA200211_H_output/1/part-m-00010
blk_1090520189 /user/hive/warehouse/prs_mc.db/dcm_prs_pdwtmdtl_s/pt_dt=2018-01-30/part-m-00004
blk_1082131961 /user/hive/warehouse/cas_mc.db/dcm_cas_nthpatel_h/end_dt=2017-12-31/000004_0
blk_1082132310 /user/hive/warehouse/cr1_mc.db/dcm_cr1_ecs_tb2045_s/pt_dt=2017-12-31/000005_0
blk_1082132604 /user/hive/warehouse/cr1_mc.db/dcm_cr1_ecs_tb2045_s/pt_dt=2017-12-31/000040_0
blk_1090521279 /user/hive/warehouse/gcm_mc.db/dcm_gcm_pndlta200211_h/end_dt=2018-01-30/000006_0
blk_1090521284 /user/hive/warehouse/gcm_mc.db/dcm_gcm_pndlta200211_h/end_dt=2018-01-30/000012_0
blk_1090521427 /user/hive/warehouse/pis_mc.db/dcm_pis_lthpcdtl_h/end_dt=2018-01-30/000080_0
blk_1090521473 /user/hive/warehouse/pis_mc.db/dcm_pis_lthpcdtl_h/end_dt=2018-01-30/000016_0
blk_1082133176 /user/hive/warehouse/cas_mc.db/dcm_cas_kffpbat_s/pt_dt=2017-12-31/part-m-00006
blk_1090522261 /user/etlhadoop/struct_data/uds_data/ECS/20180130/DCM_ECS_TB1170_S_000_input/1/ciw-20180130-hdwb1171-022_032.bin_16
blk_1090522656 /user/etlhadoop/struct_data/uds_data/ECS/20180130/DCM_ECS_TB1170_S_output/1/part-m-00007
blk_1090522747 /user/hive/warehouse/gcm_mc.db/dcm_gcm_zassure_change_detail_s/pt_dt=2018-01-31/000002_0
blk_1082134372 /user/hive/warehouse/bcs_mc.db/dcm_bcs_bthrsism_h/pt_dt=2017-12-31/part-m-00006
blk_1090523585 /user/hive/warehouse/ecs_mc.db/dcm_ecs_tb1170_s/pt_dt=2018-01-30/000002_0
blk_1090523811 /user/hive/warehouse/nae_mc.db/dcm_nae_nfpjnl_s/pt_dt=2018-01-30/part-m-00005
blk_1082135337 /user/hive/warehouse/bcs_mc.db/dcm_bcs_bthrsism_h/pt_dt=2017-12-31/part-m-00022
blk_1090524043 /user/hive/warehouse/nae_mc.db/dcm_nae_nfpjnl_s/pt_dt=2018-01-30/part-m-00016
blk_1082136206 /user/hive/warehouse/bcs_mc.db/dcm_bcs_bthrsism_h/pt_dt=2017-12-31/part-m-00038
blk_1090525355 /user/hive/warehouse/bdsp_bcas_act.db/bcs_jzcs_detail/pt_dt=2017-11-30/000006_0
blk_1090526191 /user/hive/warehouse/bdsp_bcas_act.db/bcs_jzcs_detail/pt_dt=2017-11-30/000008_0
blk_1090526995 /user/hive/warehouse/bdsp_bcas_act.db/bcs_jzcs_detail/pt_dt=2017-11-30/000014_0
blk_1082140552 /user/hive/warehouse/cc8_mc.db/m01_cc8_corp_cust_mgr/pt_dt=2017-12-31/000001_0
blk_1090529399 /user/hive/warehouse/bdsp_bcas_act.db/bcs_jzcs_middle_t/pt_dt=2017-11-30/000017_0
blk_1090529420 /user/hive/warehouse/bdsp_bcas_act.db/bcs_jzcs_middle_t/pt_dt=2017-11-30/000014_0
blk_1082141596 /user/hive/warehouse/asa_mc.db/t80_asa_bcas_agt_stat/pt_dt=2017-12-31/000032_0
blk_1082141631 /user/hive/warehouse/asa_mc.db/t80_asa_bcas_agt_stat/pt_dt=2017-12-31/000003_0
blk_1082142345 /user/hive/warehouse/sum_mc.db/c00_prod_level_overview_h/pt_dt=2017-12-31/000000_0_copy_1514441582192
blk_1090531076
/user/etlhadoop/struct_data/uds_data/GCM/20180131/DCM_GCM_DEDUN_STOP_PARA_S_000_input/1/CMA_DEDUN_STOP_PARA0111800000-011-
20180131.BIN_11_UTF
blk_1090531330 /user/hive/warehouse/gcc_mc.db/dcm_gcc_zcorp_motor_info_s/pt_dt=2018-01-31/000011_0
blk_1090531342 /user/hive/warehouse/gcc_mc.db/dcm_gcc_zcorp_motor_info_s/pt_dt=2018-01-31/000002_0
blk_1090531494
/user/etlhadoop/struct_data/uds_data/GCM/20180131/DCM_GCM_ZMORTGAGE_PROJECT_STAT_S_000_input/1/CMA_ZMORTGAGE_PROJECT_STAT0050100000-
```

2. DataNode information on the native page shows that the number of displayed DataNode nodes is 10 less than that of actual DataNode nodes.

Figure 15-24 Checking the number of DataNodes



Summary

Security is on.

Safemode is off.

14442 files and directories, 13907 blocks = 28349 total filesystem object(s).

Heap Memory used 495.63 MB of 1.99 GB Heap Memory. Max Heap Memory is 3.98 GB.

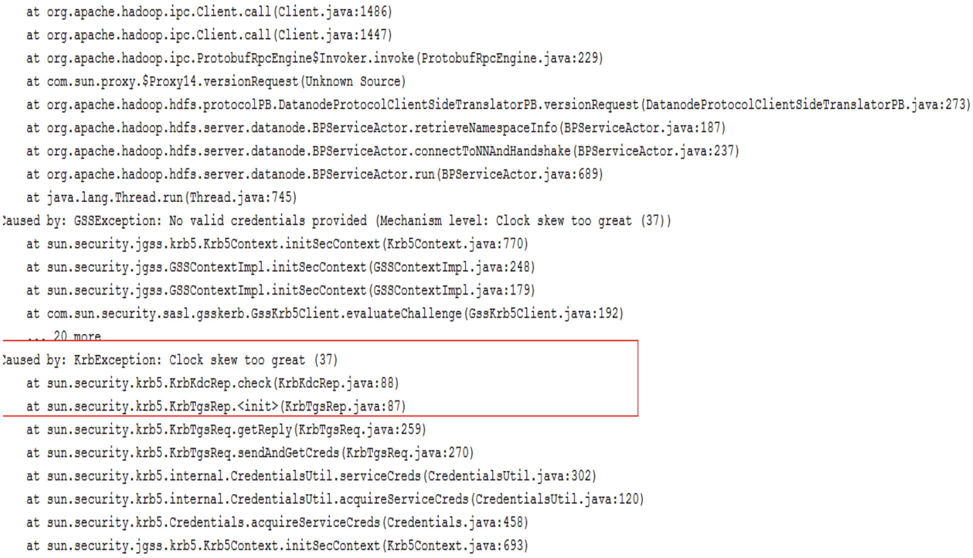
Non Heap Memory used 104.5 MB of 107.94 MB Commited Non Heap Memory. Max Non Heap Memory is 1.36 GB.

Configured Capacity:	112.09 GB
DFS Used:	15.33 GB (13.68%)
Non DFS Used:	18.56 GB
DFS Remaining:	78.2 GB (69.77%)
Block Pool Used:	15.33 GB (13.68%)
DataNodes usages% (Min/Median/Max/stdDev):	13.56% / 13.73% / 13.73% / 0.08%
Live Nodes	3 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)
Decommissioning Nodes	0

3. Check the DataNode run log file **/var/log/Bigdata/hdfs/dn/hadoop-omm-datanode-*hostname*.log**. The following error information is displayed:

   Major error information: Clock skew too great

   Figure 15-25 DateNode run log error



```
    at org.apache.hadoop.ipc.Client.call(Client.java:1486)
    at org.apache.hadoop.ipc.Client.call(Client.java:1447)
    at org.apache.hadoop.ipc.ProtobufRpcEngine$Invoker.invoke(ProtobufRpcEngine.java:229)
    at com.sun.proxy.$Proxy14.versionRequest(Unknown Source)
    at org.apache.hadoop.hdfs.protocolPB.DatanodeProtocolClientSideTranslatorPB.versionRequest(DatanodeProtocolClientSideTranslatorPB.java:273)
    at org.apache.hadoop.hdfs.server.datanode.BPServiceActor.retrieveNamespaceInfo(BPServiceActor.java:187)
    at org.apache.hadoop.hdfs.server.datanode.BPServiceActor.connectToNNAndHandshake(BPServiceActor.java:237)
    at org.apache.hadoop.hdfs.server.datanode.BPServiceActor.run(BPServiceActor.java:689)
    at java.lang.Thread.run(Thread.java:745)
Caused by: GSSException: No valid credentials provided (Mechanism level: Clock skew too great (37))
    at sun.security.jgss.krb5.Krb5Context.initSecContext(Krb5Context.java:770)
    at sun.security.jgss.GSSContextImpl.initSecContext(GSSContextImpl.java:248)
    at sun.security.jgss.GSSContextImpl.initSecContext(GSSContextImpl.java:179)
    at com.sun.security.sasl.gsskerb.GssKrb5Client.evaluateChallenge(GssKrb5Client.java:192)
    ... 20 more
Caused by: KrbException: Clock skew too great (37)
    at sun.security.krb5.KrbKdcRep.check(KrbKdcRep.java:88)
    at sun.security.krb5.KrbTgsRep.<init>(KrbTgsRep.java:87)
    at sun.security.krb5.KrbTgsReq.getReply(KrbTgsReq.java:259)
    at sun.security.krb5.KrbTgsReq.sendAndGetCreds(KrbTgsReq.java:270)
    at sun.security.krb5.internal.CredentialsUtil.serviceCreds(CredentialsUtil.java:302)
    at sun.security.krb5.internal.CredentialsUtil.acquireServiceCreds(CredentialsUtil.java:120)
    at sun.security.krb5.Credentials.acquireServiceCreds(Credentials.java:458)
    at sun.security.jgss.krb5.Krb5Context.initSecContext(Krb5Context.java:693)
```

Solution

Step 1 Change the time of the 10 DataNodes that cannot be viewed on the native page.

Step 2 On Manager, restart the DataNode instances.

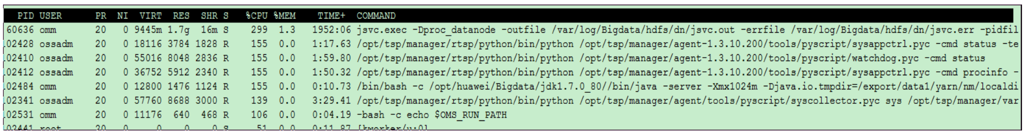
----End

15.9.10 CPU Usage of a DataNode Reaches 100% Occasionally, Causing Node Loss (SSH Connection Is Slow or Fails)

Symptom

The CPU usage of DataNodes is close to 100% occasionally, causing node loss.

Figure 15-26 DataNode CPU usage close to 100%

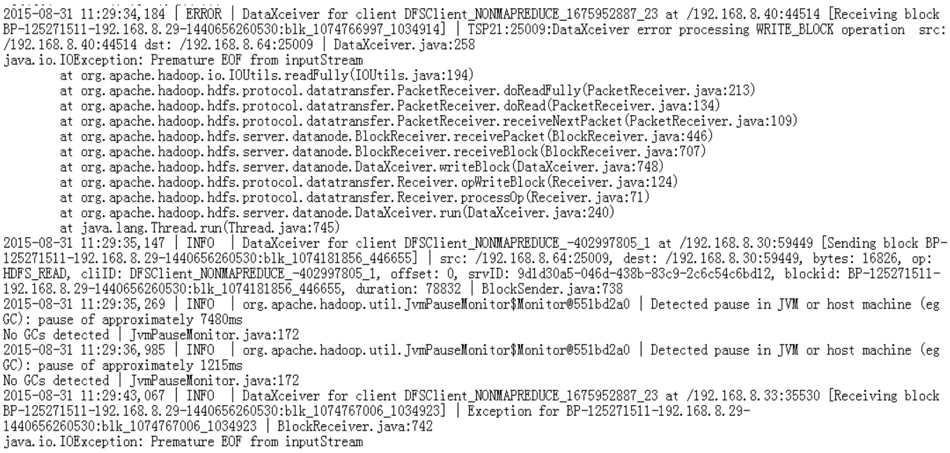


```
  PID USER      PR  NI  VIRT  RES  SHR S  %CPU %MEM    TIME+  COMMAND
60636 omm       20   0 9445m 1.7g  16m S   299  1.3  1952:06 jsvc.exec -Dproc_datanode -outfile /var/log/Bigdata/hdfs/dn/jsvc.out -errfile /var/log/Bigdata/hdfs/dn/jsvc.err -pidfil
02428 ossadm    20   0 18116 3784 1828 R   155  0.0  1:17.63 /opt/tsp/manager/rtsp/python/bin/python /opt/tsp/manager/agent-1.3.10.200/tools/pyscript/sysappctrl.pyc -cmd status -te
02410 ossadm    20   0 55016 8048 2836 R   155  0.0  1:59.80 /opt/tsp/manager/rtsp/python/bin/python /opt/tsp/manager/agent-1.3.10.200/tools/pyscript/watchdog.pyc -cmd status
02412 ossadm    20   0 36752 5912 2340 R   155  0.0  1:50.32 /opt/tsp/manager/rtsp/python/bin/python /opt/tsp/manager/agent-1.3.10.200/tools/pyscript/sysappctrl.pyc -cmd procinfo -
02484 omm       20   0 12800 1476 1124 R   155  0.0  0:10.73 /bin/bash -c /opt/huawei/Bigdata/jdk1.7.0_80//bin/java -server -Xmx1024m -Djava.io.tmpdir=/export/data1/yarn/nm/localdi
02341 ossadm    20   0 57760 8688 3000 R   139  0.0  3:29.41 /opt/tsp/manager/rtsp/python/bin/python /opt/tsp/manager/agent/tools/pyscript/syscollector.pyc sys /opt/tsp/manager/var
02531 omm       20   0 11176  640  468 R   106  0.0  0:04.19 -bash -c echo $OMS_RUN_PATH
02441 root      20   0     0    0    0 S    51  0.0  0:11.87 [kworker/u:0]
```

Cause Analysis

1. A lot of write failure logs exist on DataNodes.

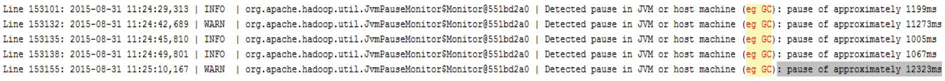
Figure 15-27 DataNode write failure log



```
2015-08-31 11:29:34,184 | ERROR | DataXceiver for client DFSClient_NONMAPREDUCE_1675952887_23 at /192.168.8.40:44514 [Receiving block
BP-125271511-192.168.8.29-1440656260530:blk_1074766997_1034914] | TSP21:25009:DataXceiver error processing WRITE_BLOCK operation  src:
/192.168.8.40:44514 dst: /192.168.8.64:25009 | DataXceiver.java:258
java.io.IOException: Premature EOF from inputStream
        at org.apache.hadoop.io.IOUtils.readFully(IOUtils.java:194)
        at org.apache.hadoop.hdfs.protocol.datatransfer.PacketReceiver.doReadFully(PacketReceiver.java:213)
        at org.apache.hadoop.hdfs.protocol.datatransfer.PacketReceiver.doRead(PacketReceiver.java:134)
        at org.apache.hadoop.hdfs.protocol.datatransfer.PacketReceiver.receiveNextPacket(PacketReceiver.java:109)
        at org.apache.hadoop.hdfs.server.datanode.BlockReceiver.receivePacket(BlockReceiver.java:446)
        at org.apache.hadoop.hdfs.server.datanode.BlockReceiver.receiveBlock(BlockReceiver.java:707)
        at org.apache.hadoop.hdfs.server.datanode.DataXceiver.writeBlock(DataXceiver.java:748)
        at org.apache.hadoop.hdfs.protocol.datatransfer.Receiver.opWriteBlock(Receiver.java:124)
        at org.apache.hadoop.hdfs.protocol.datatransfer.Receiver.processOp(Receiver.java:71)
        at org.apache.hadoop.hdfs.server.datanode.DataXceiver.run(DataXceiver.java:240)
        at java.lang.Thread.run(Thread.java:745)
2015-08-31 11:29:35,147 | INFO  | DataXceiver for client DFSClient_NONMAPREDUCE_-402997805_1 at /192.168.8.30:59449 [Sending block BP-
125271511-192.168.8.29-1440656260530:blk_1074181856_446655] | src: /192.168.8.64:25009, dest: /192.168.8.30:59449, bytes: 16826, op:
HDFS_READ, cliID: DFSClient_NONMAPREDUCE_-402997805_1, offset: 0, srvID: 9d1d30a5-046d-438b-83c9-2c6c54c6bd12, blockid: BP-125271511-
192.168.8.29-1440656260530:blk_1074181856_446655, duration: 78832 | BlockSender.java:738
2015-08-31 11:29:35,269 | INFO  | org.apache.hadoop.util.JvmPauseMonitor$Monitor@551bd2a0 | Detected pause in JVM or host machine (eg
GC): pause of approximately 7480ms
No GCs detected | JvmPauseMonitor.java:172
2015-08-31 11:29:36,985 | INFO  | org.apache.hadoop.util.JvmPauseMonitor$Monitor@551bd2a0 | Detected pause in JVM or host machine (eg
GC): pause of approximately 1215ms
No GCs detected | JvmPauseMonitor.java:172
2015-08-31 11:29:43,067 | INFO  | DataXceiver for client DFSClient_NONMAPREDUCE_1675952887_23 at /192.168.8.33:35530 [Receiving block
BP-125271511-192.168.8.29-1440656260530:blk_1074767006_1034923] | Exception for BP-125271511-192.168.8.29-
1440656260530:blk_1074767006_1034923 | BlockReceiver.java:742
java.io.IOException: Premature EOF from inputStream
```

2. A large number of files are written in a short time, causing insufficient DataNode memory.

Figure 15-28 Insufficient DataNode memory



```
Line 153101: 2015-08-31 11:24:29,313 | INFO  | org.apache.hadoop.util.JvmPauseMonitor$Monitor@551bd2a0 | Detected pause in JVM or host machine (eg GC): pause of approximately 1199ms
Line 153132: 2015-08-31 11:24:42,689 | WARN  | org.apache.hadoop.util.JvmPauseMonitor$Monitor@551bd2a0 | Detected pause in JVM or host machine (eg GC): pause of approximately 11273ms
Line 153135: 2015-08-31 11:24:45,810 | INFO  | org.apache.hadoop.util.JvmPauseMonitor$Monitor@551bd2a0 | Detected pause in JVM or host machine (eg GC): pause of approximately 1005ms
Line 153138: 2015-08-31 11:24:49,801 | INFO  | org.apache.hadoop.util.JvmPauseMonitor$Monitor@551bd2a0 | Detected pause in JVM or host machine (eg GC): pause of approximately 1067ms
Line 153155: 2015-08-31 11:25:10,167 | WARN  | org.apache.hadoop.util.JvmPauseMonitor$Monitor@551bd2a0 | Detected pause in JVM or host machine (eg GC): pause of approximately 12323ms
```

Solution

Step 1 Check DataNode memory configuration and whether the remaining server memory is sufficient.

Step 2 Increase DataNode memory and restart the DataNode.

----End

15.9.11 Manually Performing Checkpoints When a NameNode Is Faulty for a Long Time

Symptom

If the standby NameNode is faulty for a long time, a large amount of editlog will be accumulated. In this case, if the HDFS or active NameNode is restarted, the active NameNode reads a large amount of unmerged editlog. As a result, the HDFS or active NameNode takes a long time to restart and even fails to restart.

Cause Analysis

The standby NameNode periodically combines editlog files and generates the fsimage file. This process is called checkpoint. After the fsimage file is generated, the standby NameNode transfers it to the active NameNode.

NOTE

As the standby NameNode periodically combines editlog files, it cannot combine them when it becomes abnormal. As a result, the active NameNode needs to load many editlog files during its next startup, which occupies much memory and takes a long time.

The period of metadata combination is determined by the following parameters. If the NameNode runs for 30 minutes or one million counts of operations are performed on HDFS, the checkpoint is implemented.

- **dfs.namenode.checkpoint.period**: specifies the checkpoint period. The default value is **1800s**.
- dfs.namenode.checkpoint.txns: specifies the times of operations for triggering the checkpoint execution. The default value is **1000000**.

Solution

Before restarting the HDFS or active NameNode, perform checkpoint manually to merge metadata of the active NameNode.

Step 1 Stop workloads.

Step 2 Obtain the hostname of the active NameNode.

Step 3 Run the following commands on the client:

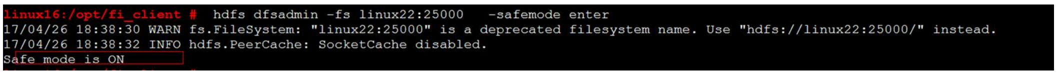
source /opt/client/bigdata_env

kinit *Component user*

Note: Replace **/opt/client** with the actual installation path of the client.

Step 4 Run the following command to enable the safe mode for the active NameNode (replace **linux22** with the hostname of the active NameNode):

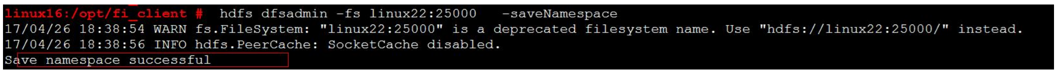
hdfs dfsadmin -fs linux22:25000 -safemode enter



```
linux16:/opt/fi_client #  hdfs dfsadmin -fs linux22:25000   -safemode enter
17/04/26 18:38:30 WARN fs.FileSystem: "linux22:25000" is a deprecated filesystem name. Use "hdfs://linux22:25000/" instead.
17/04/26 18:38:32 INFO hdfs.PeerCache: SocketCache disabled.
Safe mode is ON
```

Step 5 Run the following command to merge editlog on the active NameNode:

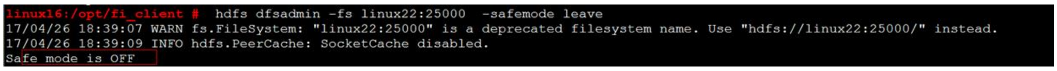
hdfs dfsadmin -fs linux22:25000 -saveNamespace



```
linux16:/opt/fi_client #  hdfs dfsadmin -fs linux22:25000   -saveNamespace
17/04/26 18:38:54 WARN fs.FileSystem: "linux22:25000" is a deprecated filesystem name. Use "hdfs://linux22:25000/" instead.
17/04/26 18:38:56 INFO hdfs.PeerCache: SocketCache disabled.
Save namespace successful
```

Step 6 Run the following command to make the active NameNode exit the safe mode:

hdfs dfsadmin -fs linux22:25000 -safemode leave

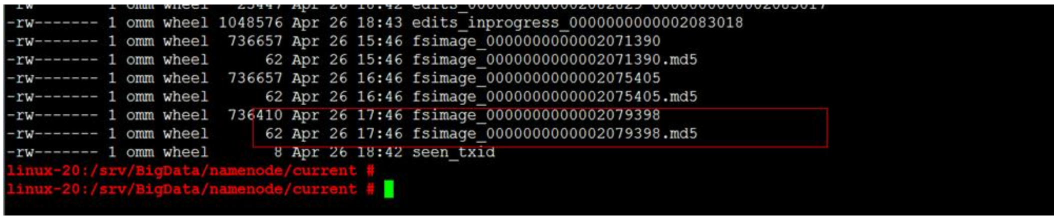


```
linux16:/opt/fi_client #  hdfs dfsadmin -fs linux22:25000  -safemode leave
17/04/26 18:39:07 WARN fs.FileSystem: "linux22:25000" is a deprecated filesystem name. Use "hdfs://linux22:25000/" instead.
17/04/26 18:39:09 INFO hdfs.PeerCache: SocketCache disabled.
Safe mode is OFF
```

Step 7 Check whether the combination is complete.

cd /srv/BigData/namenode/current

Check whether the time of the first generated fsimage is the current time. If yes, the combination is complete.



```
-rw------- 1 omm wheel 1048576 Apr 26 18:43 edits_inprogress_0000000000002083018
-rw------- 1 omm wheel  736657 Apr 26 15:46 fsimage_0000000000002071390
-rw------- 1 omm wheel      62 Apr 26 15:46 fsimage_0000000000002071390.md5
-rw------- 1 omm wheel  736657 Apr 26 16:46 fsimage_0000000000002075405
-rw------- 1 omm wheel      62 Apr 26 16:46 fsimage_0000000000002075405.md5
-rw------- 1 omm wheel  736410 Apr 26 17:46 fsimage_0000000000002079398
-rw------- 1 omm wheel      62 Apr 26 17:46 fsimage_0000000000002079398.md5
-rw------- 1 omm wheel       8 Apr 26 18:42 seen_txid
linux-20:/srv/BigData/namenode/current #
linux-20:/srv/BigData/namenode/current #
```

----End

15.9.12 Common File Read/Write Faults

Symptom

When a user performs a write operation on HDFS, the message "Failed to place enough replicas:expected..." is displayed.

Cause Analysis

- The data receiver of the DataNode is unavailable.

  The DataNode log is as follows:

  ```
  2016-03-17 18:51:44,721 | WARN |
  org.apache.hadoop.hdfs.server.datanode.DataXceiverServer@5386659f |
  hadoopc1h2:25009:DataXceiverServer: | DataXceiverServer.java:158
  java.io.IOException: Xceiver count 4097 exceeds the limit of concurrent xcievers: 4096
  at org.apache.hadoop.hdfs.server.datanode.DataXceiverServer.run(DataXceiverServer.java:140)
  at java.lang.Thread.run(Thread.java:745)
  ```

- The disk space configured for the DataNode is insufficient.
- DataNode heartbeats are delayed.

Solution

- If the DataNode data receiver is unavailable, add the value of the HDFS parameter **dfs.datanode.max.transfer.threads** on Manager.
- If disk space or CPU resources are insufficient, add DataNodes or ensure that disk space and CPU resources are available.
- If the network is faulty, ensure that the network is available.

15.9.13 Maximum Number of File Handles Is Set to a Too Small Value, Causing File Reading and Writing Exceptions

Symptom

The maximum number of file handles is set to a too small value, causing insufficient file handles. Writing files to HDFS is slow or file writing fails.

Cause Analysis

1. The DataNode log **/var/log/Bigdata/hdfs/dn/hadoop-omm-datanode-XXX.log** contains exception information "java.io.IOException: Too many open files."

   ```
   2016-05-19 17:18:59,126 | WARN  |
   org.apache.hadoop.hdfs.server.datanode.DataXceiverServer@142ff9fa |
   YSDN12:25009:DataXceiverServer:  |
   ```

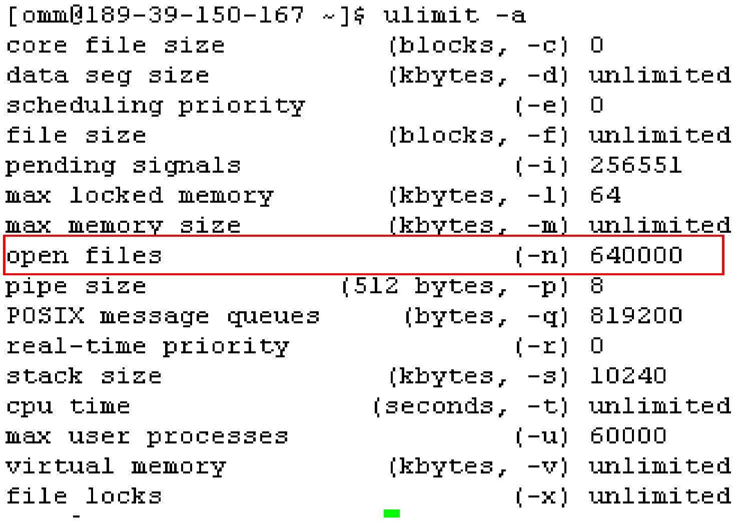
```
org.apache.hadoop.hdfs.server.datanode.DataXceiverServer.run(DataXceiverServer.java:160)
java.io.IOException: Too many open files
at sun.nio.ch.ServerSocketChannelImpl.accept0(Native Method)
at sun.nio.ch.ServerSocketChannelImpl.accept(ServerSocketChannelImpl.java:241)
at sun.nio.ch.ServerSocketAdaptor.accept(ServerSocketAdaptor.java:100)
at org.apache.hadoop.hdfs.net.TcpPeerServer.accept(TcpPeerServer.java:134)
at org.apache.hadoop.hdfs.server.datanode.DataXceiverServer.run(DataXceiverServer.java:137)
at java.lang.Thread.run(Thread.java:745)
```

2. The error indicates insufficient file handles. File handles cannot be opened and data is written to other DataNodes. As a result, writing files is slow or fails.

Solution

Step 1 Run the **ulimit -a** command to check the maximum number of file handles set for the involved node. If the value is small, change it to **640000**.

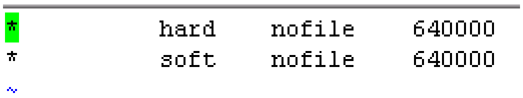
Figure 15-29 Check the number of file handles.



```
[omm@189-39-150-167 ~]$ ulimit -a
core file size          (blocks, -c) 0
data seg size           (kbytes, -d) unlimited
scheduling priority             (-e) 0
file size               (blocks, -f) unlimited
pending signals                 (-i) 256551
max locked memory       (kbytes, -l) 64
max memory size         (kbytes, -m) unlimited
open files                      (-n) 640000
pipe size            (512 bytes, -p) 8
POSIX message queues     (bytes, -q) 819200
real-time priority              (-r) 0
stack size              (kbytes, -s) 10240
cpu time               (seconds, -t) unlimited
max user processes              (-u) 60000
virtual memory          (kbytes, -v) unlimited
file locks                      (-x) unlimited
```

Step 2 Run the **vi /etc/security/limits.d/90-nofile.conf** command to edit this file. Set the number of file handles to **64000**. If the file does not exist, create one and modify the file as follows:

Figure 15-30 Changing the number of file handles



```
*          hard    nofile    640000
*          soft    nofile    640000
~
```

Step 3 Open another terminal. Run the **ulimit -a** command to check whether the modification is successful. If the modification fails, perform the preceding operations again.

Step 4 Restart the DataNode instance on Manager.

----End

15.9.14 A Client File Fails to Be Closed After Data Writing

Symptom

A client file fails to be closed after data is written to the file. A message is displayed indicating that the data block does not have enough replicas.

Client log:

```
2015-05-27 19:00:52.811 [pool-2-thread-3] ERROR: /tsp/nedata/collect/UGW/ugwufdr/
20150527/10/6_20150527105000_20150527105500_SR5S14_1432723806338_128_11.pkg.tmp143272380633
8 close hdfs sequence file fail (SequenceFileInfoChannel.java:444)
java.io.IOException: Unable to close file because the last block does not have enough number of replicas.
at org.apache.hadoop.hdfs.DFSOutputStream.completeFile(DFSOutputStream.java:2160)
at org.apache.hadoop.hdfs.DFSOutputStream.close(DFSOutputStream.java:2128)
at org.apache.hadoop.fs.FSDataOutputStream$PositionCache.close(FSDataOutputStream.java:70)
at org.apache.hadoop.fs.FSDataOutputStream.close(FSDataOutputStream.java:103)
at com.xxx.pai.collect2.stream.SequenceFileInfoChannel.close(SequenceFileInfoChannel.java:433)
at com.xxx.pai.collect2.stream.SequenceFileWriterToolChannel
$FileCloseTask.call(SequenceFileWriterToolChannel.java:804)
at com.xxx.pai.collect2.stream.SequenceFileWriterToolChannel
$FileCloseTask.call(SequenceFileWriterToolChannel.java:792)
at java.util.concurrent.FutureTask.run(FutureTask.java:262)
at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1145)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:615)
at java.lang.Thread.run(Thread.java:745)
```

Cause Analysis

1. The HDFS client starts to write blocks.

   For example, the HDFS client starts to write **/20150527/10/6_20150527105000_20150527105500_SR5S14_1432723806338_128_11.pkg.tmp1432723806338 at 2015-05-27 18:50:24,232**. The allocated block is **blk_1099105501_25370893**:

   ```
   2015-05-27 18:50:24,232 | INFO  | IPC Server handler 30 on 25000 | BLOCK* allocateBlock: /
   20150527/10/6_20150527105000_20150527105500_SR5S14_1432723806338_128_11.pkg.tmp1432723
   806338. BP-1803470917-192.168.57.33-1428597734132
   blk_1099105501_25370893{blockUCState=UNDER_CONSTRUCTION, primaryNodeIndex=-1,
   replicas=[ReplicaUnderConstruction[[DISK]DS-b2d7b7d0-f410-4958-8eba-6deecbca2f87:NORMAL|
   RBW], ReplicaUnderConstruction[[DISK]DS-76bd80e7-ad58-49c6-bf2c-03f91caf750f:NORMAL|RBW]]}
   | org.apache.hadoop.hdfs.server.namenode.FSNamesystem.saveAllocatedBlock(FSNamesystem.java:
   3166)
   ```

2. After the writing is complete, the HDFS client invokes **fsync**:

   ```
   2015-05-27 19:00:22,717 | INFO  | IPC Server handler 22 on 25000 | BLOCK* fsync:
   20150527/10/6_20150527105000_20150527105500_SR5S14_1432723806338_128_11.pkg.tmp1432723
   806338 for DFSClient_NONMAPREDUCE_-120525246_15 |
   org.apache.hadoop.hdfs.server.namenode.FSNamesystem.fsync(FSNamesystem.java:3805)
   ```

3. The HDFS client invokes **close** to close the file. After receiving the close request from the client, the NameNode uses the checkFileProgress function to check the completion status of the last block and closes the file only when enough DataNodes report that the last block is complete:

   ```
   2015-05-27 19:00:27,603 | INFO  | IPC Server handler 44 on 25000 | BLOCK* checkFileProgress:
   blk_1099105501_25370893{blockUCState=COMMITTED, primaryNodeIndex=-1,
   replicas=[ReplicaUnderConstruction[[DISK]DS-ef5fd3c9-5088-4813-ae9a-34a0714ec3a3:NORMAL|
   RBW], ReplicaUnderConstruction[[DISK]DS-f863e30f-ce5b-48cc-9cca-72f64c558adc:NORMAL|RBW]]}
   has not reached minimal replication 1 |
   org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkFileProgress(FSNamesystem.java:3197)
   2015-05-27 19:00:28,005 | INFO  | IPC Server handler 45 on 25000 | BLOCK* checkFileProgress:
   blk_1099105501_25370893{blockUCState=COMMITTED, primaryNodeIndex=-1,
   replicas=[ReplicaUnderConstruction[[DISK]DS-ef5fd3c9-5088-4813-ae9a-34a0714ec3a3:NORMAL|
   RBW], ReplicaUnderConstruction[[DISK]DS-f863e30f-ce5b-48cc-9cca-72f64c558adc:NORMAL|RBW]]}
   has not reached minimal replication 1 |
   ```

```
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkFileProgress(FSNamesystem.java:3197)
2015-05-27 19:00:28,806 | INFO  | IPC Server handler 63 on 25000 | BLOCK* checkFileProgress:
blk_1099105501_25370893{blockUCState=COMMITTED, primaryNodeIndex=-1,
replicas=[ReplicaUnderConstruction[[DISK]DS-ef5fd3c9-5088-4813-ae9a-34a0714ec3a3:NORMAL|
RBW], ReplicaUnderConstruction[[DISK]DS-f863e30f-ce5b-48cc-9cca-72f64c558adc:NORMAL|RBW]]}
has not reached minimal replication 1 |
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkFileProgress(FSNamesystem.java:3197)
2015-05-27 19:00:30,408 | INFO  | IPC Server handler 43 on 25000 | BLOCK* checkFileProgress:
blk_1099105501_25370893{blockUCState=COMMITTED, primaryNodeIndex=-1,
replicas=[ReplicaUnderConstruction[[DISK]DS-ef5fd3c9-5088-4813-ae9a-34a0714ec3a3:NORMAL|
RBW], ReplicaUnderConstruction[[DISK]DS-f863e30f-ce5b-48cc-9cca-72f64c558adc:NORMAL|RBW]]}
has not reached minimal replication 1 |
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkFileProgress(FSNamesystem.java:3197)
2015-05-27 19:00:33,610 | INFO  | IPC Server handler 37 on 25000 | BLOCK* checkFileProgress:
blk_1099105501_25370893{blockUCState=COMMITTED, primaryNodeIndex=-1,
replicas=[ReplicaUnderConstruction[[DISK]DS-ef5fd3c9-5088-4813-ae9a-34a0714ec3a3:NORMAL|
RBW], ReplicaUnderConstruction[[DISK]DS-f863e30f-ce5b-48cc-9cca-72f64c558adc:NORMAL|RBW]]}
has not reached minimal replication 1 |
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkFileProgress(FSNamesystem.java:3197)
2015-05-27 19:00:40,011 | INFO  | IPC Server handler 37 on 25000 | BLOCK* checkFileProgress:
blk_1099105501_25370893{blockUCState=COMMITTED, primaryNodeIndex=-1,
replicas=[ReplicaUnderConstruction[[DISK]DS-ef5fd3c9-5088-4813-ae9a-34a0714ec3a3:NORMAL|
RBW], ReplicaUnderConstruction[[DISK]DS-f863e30f-ce5b-48cc-9cca-72f64c558adc:NORMAL|RBW]]}
has not reached minimal replication 1 |
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkFileProgress(FSNamesystem.java:3197)
```

4. The NameNode prints **CheckFileProgress** multiple times because the HDFS client retries to close the file for several times. The file closing fails because the block status is not complete. The number of retries is determined by the **dfs.client.block.write.locateFollowingBlock.retries** parameter. The default value is **5**. Therefore, **CheckFileProgress** is printed six times in the NameNode log.
5. After 0.5 seconds, the DataNodes report that the block has been successfully written.

```
2015-05-27 19:00:40,608 | INFO  | IPC Server handler 60 on 25000 | BLOCK* addStoredBlock:
blockMap updated: 192.168.10.21:25009 is added to
blk_1099105501_25370893{blockUCState=COMMITTED, primaryNodeIndex=-1,
replicas=[ReplicaUnderConstruction[[DISK]DS-ef5fd3c9-5088-4813-ae9a-34a0714ec3a3:NORMAL|
RBW], ReplicaUnderConstruction[[DISK]DS-f863e30f-ce5b-48cc-9cca-72f64c558adc:NORMAL|RBW]]}
size 11837530 |
org.apache.hadoop.hdfs.server.blockmanagement.BlockManager.logAddStoredBlock(BlockManager.java
:2393)
2015-05-27 19:00:48,297 | INFO  | IPC Server handler 37 on 25000 | BLOCK* addStoredBlock:
blockMap updated: 192.168.10.10:25009 is added to blk_1099105501_25370893 size 11837530 |
org.apache.hadoop.hdfs.server.blockmanagement.BlockManager.logAddStoredBlock(BlockManager.java
:2393)
```

6. The block write success notification is delayed because of network bottlenecks or CPU bottlenecks.
7. If close is invoked again or the number of file closing retries increases, a closing success message will be displayed. You are advised to increase the value of **dfs.client.block.write.locateFollowingBlock.retries**. The default parameter value is 5 and retry intervals are 400 ms, 800 ms, 1600 ms, 3200 ms, 6400 ms, and 12800 ms. Therefore, the result of the close function can be returned after a maximum of 25.2 seconds.

Solution

Step 1 Solution:

Set the value of **dfs.client.block.write.locateFollowingBlock.retries** to **6**. The retry intervals are 400 ms, 800 ms, 1600 ms, 3200 ms, 6400 ms, and 12800 ms.

Therefore, the result of the close function can be returned after a maximum of 50.8 seconds.

----End

Remarks

Generally, this fault occurs when the cluster workload is heavy. Adjusting the parameter can only temporarily avoid the fault. You are advised to reduce the cluster workload, for example, do not allocate all CPU resources to MapReduce.

15.9.15 File Fails to Be Uploaded to HDFS Due to File Errors

Symptom

The **hadoop dfs -put** command is used to copy local files to HDFS.

After some files are uploaded, an error occurs. The size of the temporary files no long changes on the native NameNode page.

Cause Analysis

1. Check the NameNode log **/var/log/Bigdata/hdfs/nn/hadoop-omm-namenode-*hostname*.log**. It is found that the file is being written until a failure occurs.
```
2015-07-13 10:05:07,847 | WARN  | org.apache.hadoop.hdfs.server.namenode.LeaseManager
$Monitor@36fea922 | DIR* NameSystem.internalReleaseLease: Failed to release lease for file /hive/
order/OS_ORDER._8.txt._COPYING_. Committed blocks are waiting to be minimally replicated. Try
again later. | FSNamesystem.java:3936
2015-07-13 10:05:07,847 | ERROR | org.apache.hadoop.hdfs.server.namenode.LeaseManager
$Monitor@36fea922 | Cannot release the path /hive/order/OS_ORDER._8.txt._COPYING_ in the lease
[Lease.  Holder: DFSClient_NONMAPREDUCE_-1872896146_1, pendingcreates: 1] | LeaseManager.java:
459
org.apache.hadoop.hdfs.protocol.AlreadyBeingCreatedException: DIR*
NameSystem.internalReleaseLease: Failed to release lease for file /hive/order/
OS_ORDER._8.txt._COPYING_. Committed blocks are waiting to be minimally replicated. Try again
later.
at FSNamesystem.internalReleaseLease(FSNamesystem.java:3937)
```
2. Root cause: The uploaded files are damaged.
3. Verification: The cp or scp operation fails to be performed for the copied files. Therefore, the files are damaged.

Solution

Step 1 Upload normal files.

----End

15.9.16 After dfs.blocksize Is Configured and Data Is Put, Block Size Remains Unchanged

Symptom

After **dfs.blocksize** is set to **268435456** on the interface and data is put, the original block size keeps unchanged.

Cause Analysis

The **dfs.blocksize** value in the **hdfs-site.xml** file of the client is not changed, and the value prevails.

Solution

Step 1 Ensure that the **dfs.blocksize** value is a multiple of 512.

Step 2 Download a client or modify the client configuration.

Step 3 **dfs.blocksize** is configured on the client and is subject to the client. Otherwise, the value configured on the server prevails.

----End

15.9.17 Failed to Read Files, and "FileNotFoundException" Is Displayed

Symptom

In MapReduce tasks, all Map tasks are successfully executed, but Reduce tasks fail. The error message "FileNotFoundException...No lease on...File does not exist" is displayed in the logs.

```
Error: org.apache.hadoop.ipc.RemoteException(java.io.FileNotFoundException): No lease on /user/sparkhive/
warehouse/daas/dsp/output/_temporary/1/_temporary/attempt_1479799053892_17075_r_000007_0/part-
r-00007 (inode 6501287): File does not exist. Holder
DFSClient_attempt_1479799053892_17075_r_000007_0_-1463597952_1 does not have any open files.
at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkLease(FSNamesystem.java:3350)
at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.completeFileInternal(FSNamesystem.java:3442)
at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.completeFile(FSNamesystem.java:3409)
at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.complete(NameNodeRpcServer.java:789)
```

Cause Analysis

"FileNotFoundException...No lease on...File does not exist" indicates that the file is deleted during the operation.

1. Search for the file name in the NameNode audit log of HDFS (**/var/log/Bigdata/audit/hdfs/nn/hdfs-audit-namenode.log** of the active NameNode) to confirm the creation time of the file.
2. Search the NameNode audit logs that are generated within the time range from the file creation to the time of exception occurrence and determine whether the file is deleted or moved to another directory.
3. If the file is not deleted or moved, the parent directory of the file may be deleted or moved. You need to search the upper-layer directory. In this example, the parent directory of the file's parent directory is deleted.
   ```
   2017-05-31 02:04:08,286 | INFO  | IPC Server handler 30 on 25000 | allowed=true
   ugi=appUser@HADOOP.COM (auth:TOKEN) ip=/192.168.1.22     cmd=delete      src=/user/sparkhive/
   warehouse/daas/dsp/output/_temporary      dst=null    perm=null proto=rpc | FSNamesystem.java:8189
   ```

NOTE

- The preceding log indicates that the **appUser** user of the 192.168.1.22 node deletes **/user/sparkhive/warehouse/daas/dsp/output/_temporary**.
- Run the **zgrep "***file name***" *.zip** command to search for the contents of the .zip package.

Solution

Step 1 Check the service to find out why the file or the parent directory of the file is deleted.

----End

15.9.18 Failed to Write Files to HDFS, and "item limit of / is exceeded" Is Displayed

Symptom

The client or upper-layer component logs indicate that a file fails to be written to a directory on HDFS. The error information is as follows:

The directory item limit of /tmp is exceeded: limit=5 items=5.

Cause Analysis

1. The run log file **/var/log/Bigdata/hdfs/nn/hadoop-omm-namenode-XXX.log** of the client or NameNode contains error information "The directory item limit of /tmp is exceeded:." The error message indicates that the number of files in the **/tmp** directory exceeds 1048576.
```
2018-03-14 11:18:21,625 | WARN  | IPC Server handler 62 on 25000 | DIR* NameSystem.startFile: /tmp/
test.txt The directory item limit of /tmp is exceeded: limit=1048576 items=1048577 |
FSNamesystem.java:2334
```
2. The **dfs.namenode.fs-limits.max-directory-items** parameter specifies the maximum number of directories or files that are not in recursion relationship in a single directory. The default value is **1048576**. The value ranges from 1 to 6400000.

Solution

Step 1 Check whether it is normal that the directory contains more than one million files that are not in recursion relationship. If it is normal, increase the value of the HDFS parameter **dfs.namenode.fs-limits.max-directory-items** and restart the HDFS NameNode for the modification to take effect.

Step 2 If it is abnormal, delete unnecessary files.

----End

15.9.19 Adjusting the Log Level of the Shell Client

- **Temporary adjustment**: After the Shell client window is closed, the log is restored to the default value.

a. Run the **export HADOOP_ROOT_LOGGER** command to adjust the log level of the client.

b. Run the **export HADOOP_ROOT_LOGGER=***log level***,console** command to adjust the log level of the Shell client.

Run the **export HADOOP_ROOT_LOGGER=DEBUG,console** command to adjust the log level to **Debug**.

Run the **export HADOOP_ROOT_LOGGER=ERROR,console** command to adjust the log level to **Error**.

- **Permanent adjustment**

a. Add **export HADOOP_ROOT_LOGGER=***log level***,console** to the HDFS client's environment variable configuration file **/opt/client/HDFS/component_env** (replace **/opt/client** with the actual client path).

b. Run the **source /opt/client/bigdata_env** command.

c. Run the command on the client again.

15.9.20 File Read Fails, and "No common protection layer" Is Displayed

Symptom

HDFS fails to be operated on the Shell client or other clients, and the error message "No common protection layer between client and server" is displayed.

Running any **hadoop** command, such as **hadoop fs -ls /**, on a node outside the cluster fails. The bottom-layer error message is displayed stating "No common protection layer between client and server."

```
2017-05-13 19:14:19,060 | ERROR | [pool-1-thread-1] |  Server startup failure  | 
org.apache.sqoop.core.SqoopServer.initializeServer(SqoopServer.java:69)
org.apache.sqoop.common.SqoopException: MAPRED_EXEC_0028:Failed to operate HDFS - Failed to get the 
file /user/loader/etl_dirty_data_dir status
        at org.apache.sqoop.job.mr.HDFSClient.fileExist(HDFSClient.java:85)
...
        at java.lang.Thread.run(Thread.java:745)
Caused by: java.io.IOException: Failed on local exception: java.io.IOException: Couldn't setup connection for 
loader/hadoop@HADOOP.COM to loader37/10.162.0.37:25000; Host Details : local host is: 
"loader37/10.162.0.37"; destination host is: "loader37":25000;
        at org.apache.hadoop.net.NetUtils.wrapException(NetUtils.java:776)
...
        ... 10 more
Caused by: java.io.IOException: Couldn't setup connection for loader/hadoop@HADOOP.COM to 
loader37/10.162.0.37:25000
        at org.apache.hadoop.ipc.Client$Connection$1.run(Client.java:674
        ... 28 more
Caused by: javax.security.sasl.SaslException: No common protection layer between client and server
        at com.sun.security.sasl.gsskerb.GssKrb5Client.doFinalHandshake(GssKrb5Client.java:251)
...
        at org.apache.hadoop.ipc.Client$Connection.setupIOstreams(Client.java:720)
```

Cause Analysis

1. The RPC protocol is used for data transmission between the client and server of HDFS. The protocol has multiple encryption modes and the hadoop.rpc.protection parameter specifies the mode to use.

2. If the value of the **hadoop.rpc.protection** parameter on the client is different from that on the server, the "No common protection layer between client and server" error is reported.

   NOTE

   hadoop.rpc.protection indicates that data can be transmitted between nodes in any of the following modes:

   - **privacy**: Data is transmitted after authentication and encryption. This mode reduces the performance.
   - **authentication**: Data is transmitted after authentication without encryption. This mode ensures performance but has security risks.
   - **integrity**: Data is transmitted without encryption or authentication. To ensure data security, exercise caution when using this mode.

Solution

Step 1 Download the client again. If the client is an application, update the configuration file in the application.

----End

15.9.21 Failed to Write Files Because the HDFS Directory Quota Is Insufficient

Symptom

After quota is set for a directory, writing files to the directory fails. The "The DiskSpace quota of /tmp/tquota2 is exceeded" error message is displayed.

```
[omm@189-39-150-115 client]$ hdfs dfs -put switchuser.py  /tmp/tquota2
put: The DiskSpace quota of /tmp/tquota2 is exceeded: quota = 157286400 B = 150 MB but diskspace
consumed = 402653184 B = 384 MB
```

Possible Causes

The remaining space configured for the directory is less than the space required for writing files.

Cause Analysis

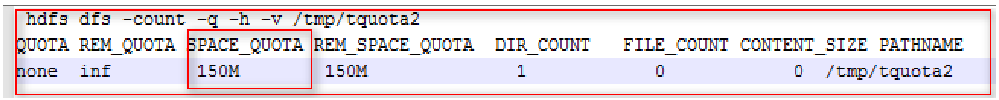
1. The HDFS supports setting the quota for a specific directory, that is, the maximum space occupied by files in a directory can be set. For example, the following command is used to set a maximum of 150 MB files to be written to the **/tmp/tquota** directory. (Space = Block size x Number of copies)

   hadoop dfsadmin -setSpaceQuota 150M /tmp/tquota2

2. Run the following command to check the configured quota for the directory. **SPACE_QUOTA** is the configured space quota, and **REM_SPACE_QUOTA** is the remaining space.

   hdfs dfs -count -q -h -v /tmp/tquota2

   Figure 15-31 Viewing the quota set for a directory



```
 hdfs dfs -count -q -h -v /tmp/tquota2
QUOTA REM_QUOTA SPACE_QUOTA REM_SPACE_QUOTA  DIR_COUNT   FILE_COUNT CONTENT_SIZE PATHNAME
none  inf        150M       150M                 1           0            0  /tmp/tquota2
```

3. Analyze logs. The following log indicates that writing the file requires 384 MB space, but the current space quota is only 150 MB. Therefore, the space is insufficient. Before a file is written, the required remaining space is as follows: Block size x Number of copies. 128 MB x 3 copies = 384 MB.

```
[omm@189-39-150-115 client]$
[omm@189-39-150-115 client]$ hdfs dfs -put switchuser.py  /tmp/tquota2
put: The DiskSpace quota of /tmp/tquota2 is exceeded: quota = 157286400 B = 150 MB but diskspace
consumed = 402653184 B = 384 MB
```

Solution

Step 1 Set a proper quota for the directory.

hadoop dfsadmin -setSpaceQuota 150G /*directory name*

Step 2 Run the following command to clear the quota:

hdfs dfsadmin -clrSpaceQuota /*directory name*

----End

15.9.22 Balancing Fails, and "Source and target differ in block-size" Is Displayed

Symptom

When the **distcp** command is executed to copy files across clusters, the message "Source and target differ in block-size." is displayed, indicating that some files fail to be copied. Use -pb to preserve block-sizes during copy. "

```
 Caused by: java.io.IOException: Check-sum mismatch between hdfs://10.180.144.7:25000/kylin/
kylin_default_instance_prod/parquet/f2e72874-f01c-45ff-b219-207f3a5b3fcb/c769cd2d-575a-4459-837b-
a19dd7b20c27/339114721280/0.parquettar and hdfs://10.180.180.194:25000/kylin/
kylin_default_instance_prod/parquet/f2e72874-f01c-45ff-
b219-207f3a5b3fcb/.distcp.tmp.attempt_1523424430246_0004_m_000019_2. Source and target differ in
block-size. Use -pb to preserve block-sizes during copy. Alternatively, skip checksum-checks altogether,
using -skipCrc. (NOTE: By skipping checksums, one runs the risk of masking data-corruption during file-
transfer.)        at
org.apache.hadoop.tools.mapred.RetriableFileCopyCommand.compareCheckSums(RetriableFileCopyComman
d.java:214)
```

Possible Causes

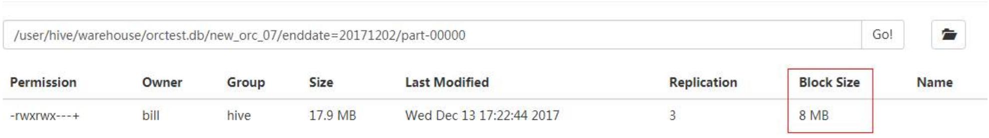
This is not a version-related problem. When you run the **distcp** command to copy files, the block size of the source file is not recorded by default. As a result, the verification fails when the block size of the source file is not 128 MB. In this case, you need to add parameter **-pb** to the **distcp** command.

Cause Analysis

1. The block size is set when data is written to HDFS. The default block size is 128 MB. The size of files written by some components or service programs may not be 128 MB, for example, 8 MB.

```
<name>dfs.blocksize</name>
<value>134217728</value>
```

Figure 15-32 Size of files written by some components or service programs



/user/hive/warehouse/orctest.db/new_orc_07/enddate=20171202/part-00000 Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rwxrwx---+	bill	hive	17.9 MB	Wed Dec 13 17:22:44 2017	3	8 MB	

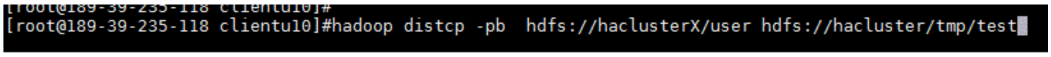
2. DistCp reads the file from a source cluster and writes it to a destination cluster. By default, the value of dfs.blocksize in the MapReduce task is used as the block size, whose default value is 128 MB.
3. After DistCp finishes writing a file, the system performs verification based on the physical size of the block. Because the block size of the file in the source cluster is different from that of the file in the destination cluster, the splitting sizes are different. As a result, the verification fails.

   For example, in the preceding file, there are three blocks (17.9/8 MB = 3 blocks) in the old cluster and one block (17.9/128 MB = 1 block) in the new cluster. Therefore, the verification fails because the physical size of the disk is divided.

Solution

Add parameter **-pb** in the **distcp** command. This parameter is used to reserve the block size when **distcp** is used to ensure that the block size of the new cluster is the same as that of the old cluster.

Figure 15-33 Size of the reserved block during **distcp** command execution



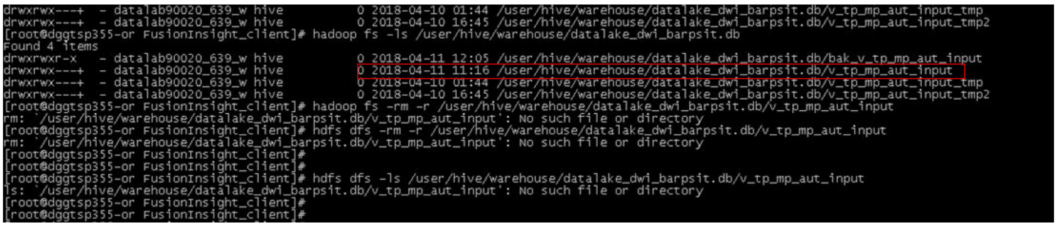
```
[root@189-39-235-118 clientu10]#
[root@189-39-235-118 clientu10]#hadoop distcp -pb  hdfs://haclusterX/user hdfs://hacluster/tmp/test
```

15.9.23 A File Fails to Be Queried or Deleted, and the File Can Be Viewed in the Parent Directory (Invisible Characters)

Symptom

A file fails to be queried or deleted using the HDFS Shell client. The file can be viewed in the parent directory.

Figure 15-34 List of files in the parent directory



```
drwxrwx---+  - datalab90020_639_w hive          0 2018-04-10 01:44 /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input_tmp
drwxrwx---+  - datalab90020_639_w hive          0 2018-04-10 16:45 /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input_tmp2
[root@dggtsp355-or FusionInsight_client]# hadoop fs -ls /user/hive/warehouse/datalake_dwi_barpsit.db
Found 4 items
drwxrwxr-x   - datalab90020_639_w hive          0 2018-04-11 12:05 /user/hive/warehouse/datalake_dwi_barpsit.db/bak_v_tp_mp_aut_input
drwxrwx---+  - datalab90020_639_w hive          0 2018-04-11 11:16 /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input
drwxrwx---+  - datalab90020_639_w hive          0 2018-04-10 01:44 /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input_tmp
drwxrwx---+  - datalab90020_639_w hive          0 2018-04-10 16:45 /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input_tmp2
[root@dggtsp355-or FusionInsight_client]# hadoop fs -rm -r /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input
rm: `/user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input': No such file or directory
[root@dggtsp355-or FusionInsight_client]# hdfs dfs -rm -r /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input
rm: `/user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input': No such file or directory
[root@dggtsp355-or FusionInsight_client]#
[root@dggtsp355-or FusionInsight_client]#
[root@dggtsp355-or FusionInsight_client]# hdfs dfs -ls /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input
ls: `/user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input': No such file or directory
[root@dggtsp355-or FusionInsight_client]#
[root@dggtsp355-or FusionInsight_client]#
```

Cause Analysis

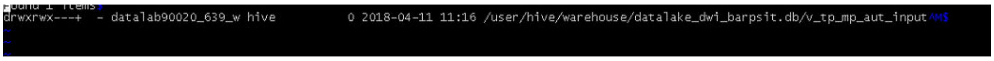
The possible cause is that invisible characters are written to the file. You can write the file name to the local text and run the **vi** command to open the file.

hdfs dfs -ls *parent directory* **> /tmp/t.txt**

vi /tmp/t.txt

Run the **:set list** command to display invisible characters in the file name. For example, the file name contains **^M**, which is invisible.

Figure 15-35 Displaying invisible characters



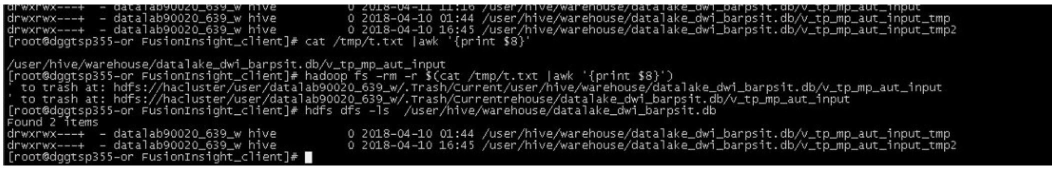
```
Found 1 items$
drwxrwx---+  - datalab90020_639_w hive          0 2018-04-11 11:16 /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input^M$
~
~
~
```

Solution

Step 1 Run the Shell command to read the file name recorded in the text. Ensure that the following command output contains the full path of the file in HDFS.

cat /tmp/t.txt |awk '{print $8}'

Figure 15-36 File path



```
drwxrwx---+  - datalab90020_639_w hive          0 2018-04-11 11:16 /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input
drwxrwx---+  - datalab90020_639_w hive          0 2018-04-10 01:44 /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input_tmp
drwxrwx---+  - datalab90020_639_w hive          0 2018-04-10 16:45 /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input_tmp2
[root@dggtsp355-or FusionInsight_client]# cat /tmp/t.txt |awk '{print $8}'

/user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input
[root@dggtsp355-or FusionInsight_client]# hadoop fs -rm -r $(cat /tmp/t.txt |awk '{print $8}')
' to trash at: hdfs://hacluster/user/datalab90020_639_w/.Trash/Current/user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input
' to trash at: hdfs://hacluster/user/datalab90020_639_w/.Trash/Currentrehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input
[root@dggtsp355-or FusionInsight_client]# hdfs dfs -ls  /user/hive/warehouse/datalake_dwi_barpsit.db
Found 2 items
drwxrwx---+  - datalab90020_639_w hive          0 2018-04-10 01:44 /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input_tmp
drwxrwx---+  - datalab90020_639_w hive          0 2018-04-10 16:45 /user/hive/warehouse/datalake_dwi_barpsit.db/v_tp_mp_aut_input_tmp2
[root@dggtsp355-or FusionInsight_client]# █
```

Step 2 Run the following command to delete the file:

hdfs dfs -rm $(cat /tmp/t.txt |awk '{print $8}')

Step 3 Verify that the file has been deleted.

hdfs dfs -ls *parent directory*

----End

15.9.24 Uneven Data Distribution Due to Non-HDFS Data Residuals

Symptom

Data distribution is uneven. A disk is full while other disks have sufficient space.

The data storage directory of HDFS DataNode is set to **/export/data1/dfs--/export/data12/dfs**. A large volume of data is stored to **/export/data1/dfs** but data is evenly distributed to other disks.

Cause Analysis

The customer's disk is reinstalled. However, a directory is not thoroughly deleted during disk uninstallation, that is, the added disk is unformatted and historical junk data remains.

Solution

Manually delete data residuals.

15.9.25 Uneven Data Distribution Due to the Client Installation on the DataNode

Symptom

Data is unevenly distributed on HDFS DataNodes. Disk usage of a node is high or even reaches 100% while disks on other nodes have sufficient idle space.

Cause Analysis

In the HDFS data replica mechanism, the first replica is stored to the local node where the client is stored. As a result, disks of the node run out while disks of other nodes have sufficient idle space.

Solution

Step 1 For the existing data unevenly distributed, run the following command to balance data:

/opt/client/HDFS/hadoop/sbin/start-balancer.sh -threshold 10

/opt/client indicates the actual client installation directory.

Step 2 For new data, install the client on the node without DataNode.

----End

15.9.26 Handling Unbalanced DataNode Disk Usage on Nodes

Symptom

The disk usage of each DataNode on a node is uneven.

Example:

```
189-39-235-71:~ # df -h
Filesystem  Size  Used Avail Use% Mounted on
/dev/xvda  360G  92G   250G  28% /
/dev/xvdb  700G  900G   200G  78% /srv/BigData/hadoop/data1
/dev/xvdc  700G  900G   200G  78% /srv/BigData/hadoop/data2
/dev/xvdd  700G  900G   200G  78% /srv/BigData/hadoop/data3
/dev/xvde  700G  900G   200G  78% /srv/BigData/hadoop/data4
/dev/xvdf  10G   900G   890G  2% /srv/BigData/hadoop/data5
189-39-235-71:~ #
```

Possible Causes

Some disks are faulty and are replaced with new ones. The new disk usage is low.

Disks are added. For example, the original four data disks are expanded to five disks.

Cause Analysis

There are two policies for writing data to Block disks on DataNodes: 1. Round Robin (default value) and 2. Preferentially writing data to the disk with the more available space.

Description of the **dfs.datanode.fsdataset.volume.choosing.policy** parameter

Possible values:

- Polling: **org.apache.hadoop.hdfs.server.datanode.fsdataset.RoundRobinVolumeChoosingPolicy**
- Preferentially writing data to the disk with more available space: **org.apache.hadoop.hdfs.server.datanode.fsdataset.AvailableSpaceVolumeChoosingPolicy**

Solution

Change the value of **dfs.datanode.fsdataset.volume.choosing.policy** to **org.apache.hadoop.hdfs.server.datanode.fsdataset.AvailableSpaceVolumeChoosingPolicy**, save the settings, and restart the affected services or instances.

In this way, the DataNode preferentially selects a node with the most available disk space to store data copies.

NOTE

- Data written to the DataNode will be preferentially written to the disk with more available disk space.
- The high usage of some disks can be relieved with the gradual deletion of aging data from the HDFS.

15.9.27 Locating Common Balance Problems

Problem 1: Lack of Permission to Execute the balance Task (Access denied).

Problem details: After the **start-balancer.sh** command is executed, the " hadoop-root-balancer-hostname.out" log displays "Access denied for user test1. Superuser privilege is required."

```
cat /opt/client/HDFS/hadoop/logs/hadoop-root-balancer-host2.out
Time Stamp               Iteration#  Bytes Already Moved  Bytes Left To Move  Bytes Being Moved
INFO: Watching file:/opt/client/HDFS/hadoop/etc/hadoop/log4j.properties for changes with interval : 60000
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.AccessControlException): Access denied
for user test1.
Superuser privilege is required
at
org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkSuperuserPrivilege(FSPermissionChecker
.java:122)
at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkSuperuserPrivilege(FSNamesystem.java:
5916)
```

Cause analysis:

The administrator account is required for executing the balance task.

Solution

- Secure version

  Perform authentication for user **hdfs** or a user in the **supergroup** group and then execute the balance task.
- General version

Run the **su - hdfs** command on the client before running the **balance** command on HDFS.

Problem 2: The balance command fails to be executed, and the /system/balancer.id file is abnormal.

Problem details:

A user starts a balance process on the HDFS client. After the process is stopped unexpectedly, the user performs the balance operation again. The operation fails.

```
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.protocol.RecoveryInProgressException): Failed
to APPEND_FILE /system/balancer.id for DFSClient because lease recovery is in progress. Try again later.
```

Cause analysis:

Generally, after the balance operation is complete in HDFS, the **/system/balancer.id** file is automatically released and the balance operation can be performed again.

In the preceding scenario, the first balance operation is stopped abnormally. Therefore, when the balance operation is performed for the second time, the **/system/balancer.id** file still exists. As a result, the **append /system/balancer.id** operation is triggered and the balance operation fails.

Solution

Method 1: After the hard lease period exceeds one hour, release the lease on the original client and perform the balance operation again.

Method 2: Delete the **/system/balancer.id** file from HDFS and perform the balance operation again.

15.9.28 HDFS Displays Insufficient Disk Space But 10% Disk Space Remains

Symptom

1. The alarm "HDFS Disk Usage Exceeds the Threshold" is reported.
2. On the HDFS page, high disk space usage is displayed.

Cause Analysis

The **dfs.datanode.du.reserved.percentage** parameter is set in HDFS, indicating the percentage of the reserved space of each disk to the total disk space. The DataNode reserves space you set for NodeManager running and computing of other components, for example, Yarn, or for upgrades.

As 10% disk space is reserved, the HDFS DataNode regards that there is no available disk space when the disk usage reaches 90%.

Solution

Step 1 Expand the HDFS DataNode disk capacity when its usage reaches 80%.

Step 2 If the disk capacity cannot be expanded in time, delete useless data in HDFS to release disk space.

----End

15.9.29 An Error Is Reported When the HDFS Client Is Installed on the Core Node in a Common Cluster

Issue

In a common cluster, an error message is displayed when a user is created on the Core node to install the client.

Symptom

In a common cluster, the following error message is displayed when a user is created on the Core node to install the client:

```
2020-03-14 19:16:17,166 WARN shortcircuit.DomainSocketFactory: error creating DomainSocket
java.net.ConnectException: connect(2) error: Permission denied when trying to connect to '/var/run/MRS-
HDFS/dn_socket'
at org.apache.hadoop.net.unix.DomainSocket.connect0(Native Method)
at org.apache.hadoop.net.unix.DomainSocket.connect(DomainSocket.java:256)
at org.apache.hadoop.hdfs.shortcircuit.DomainSocketFactory.createSocket(DomainSocketFactory.java:168)
at org.apache.hadoop.hdfs.client.impl.BlockReaderFactory.nextDomainPeer(BlockReaderFactory.java:799)
...
```

Cause Analysis

A user runs the **useradd** command to create a user. The default user group of the user does not contain the **ficommmon** user group. As a result, the preceding error is reported when the **get** command of HDFS is executed.

Procedure

Run the **usermod -a -G ficommon username** command to add the user to the **ficommon** user group.

15.9.30 Client Installed on a Node Outside the Cluster Fails to Upload Files Using hdfs

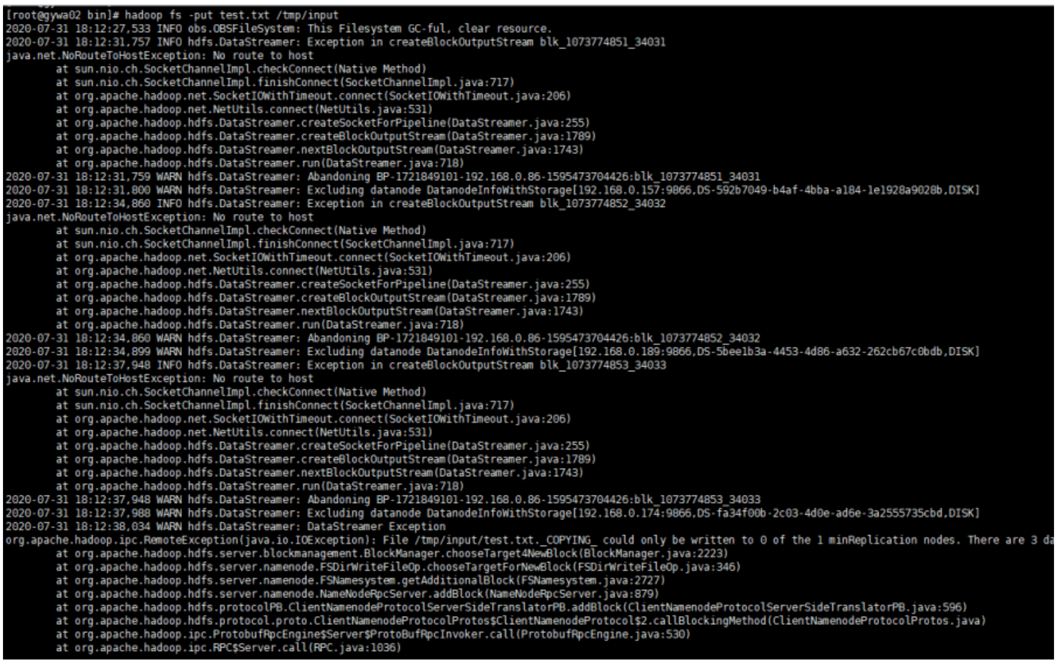
Issue

A client installed on a node outside the cluster fails to upload files using hdfs.

Symptom

After a client is installed on a cluster node and a file is uploaded using the **hdfs** command, the following error is reported:

Figure 15-37 An error is reported during file upload.



```
[root@gywa02 bin]# hadoop fs -put test.txt /tmp/input
2020-07-31 18:12:27,533 INFO obs.OBSFileSystem: This Filesystem GC-ful, clear resource.
2020-07-31 18:12:31,757 INFO hdfs.DataStreamer: Exception in createBlockOutputStream blk_1073774851_34031
java.net.NoRouteToHostException: No route to host
        at sun.nio.ch.SocketChannelImpl.checkConnect(Native Method)
        at sun.nio.ch.SocketChannelImpl.finishConnect(SocketChannelImpl.java:717)
        at org.apache.hadoop.net.SocketIOWithTimeout.connect(SocketIOWithTimeout.java:206)
        at org.apache.hadoop.net.NetUtils.connect(NetUtils.java:531)
        at org.apache.hadoop.hdfs.DataStreamer.createSocketForPipeline(DataStreamer.java:255)
        at org.apache.hadoop.hdfs.DataStreamer.createBlockOutputStream(DataStreamer.java:1789)
        at org.apache.hadoop.hdfs.DataStreamer.nextBlockOutputStream(DataStreamer.java:1743)
        at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:718)
2020-07-31 18:12:31,759 WARN hdfs.DataStreamer: Abandoning BP-1721849101-192.168.0.86-1595473704426:blk_1073774851_34031
2020-07-31 18:12:31,800 WARN hdfs.DataStreamer: Excluding datanode DatanodeInfoWithStorage[192.168.0.157:9866,DS-592b7049-b4af-4bba-a184-1e1928a9028b,DISK]
2020-07-31 18:12:34,860 INFO hdfs.DataStreamer: Exception in createBlockOutputStream blk_1073774852_34032
java.net.NoRouteToHostException: No route to host
        at sun.nio.ch.SocketChannelImpl.checkConnect(Native Method)
        at sun.nio.ch.SocketChannelImpl.finishConnect(SocketChannelImpl.java:717)
        at org.apache.hadoop.net.SocketIOWithTimeout.connect(SocketIOWithTimeout.java:206)
        at org.apache.hadoop.net.NetUtils.connect(NetUtils.java:531)
        at org.apache.hadoop.hdfs.DataStreamer.createSocketForPipeline(DataStreamer.java:255)
        at org.apache.hadoop.hdfs.DataStreamer.createBlockOutputStream(DataStreamer.java:1789)
        at org.apache.hadoop.hdfs.DataStreamer.nextBlockOutputStream(DataStreamer.java:1743)
        at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:718)
2020-07-31 18:12:34,860 WARN hdfs.DataStreamer: Abandoning BP-1721849101-192.168.0.86-1595473704426:blk_1073774852_34032
2020-07-31 18:12:34,899 WARN hdfs.DataStreamer: Excluding datanode DatanodeInfoWithStorage[192.168.0.189:9866,DS-5bee1b3a-4453-4d86-a632-262cb67c0bdb,DISK]
2020-07-31 18:12:37,948 INFO hdfs.DataStreamer: Exception in createBlockOutputStream blk_1073774853_34033
java.net.NoRouteToHostException: No route to host
        at sun.nio.ch.SocketChannelImpl.checkConnect(Native Method)
        at sun.nio.ch.SocketChannelImpl.finishConnect(SocketChannelImpl.java:717)
        at org.apache.hadoop.net.SocketIOWithTimeout.connect(SocketIOWithTimeout.java:206)
        at org.apache.hadoop.net.NetUtils.connect(NetUtils.java:531)
        at org.apache.hadoop.hdfs.DataStreamer.createSocketForPipeline(DataStreamer.java:255)
        at org.apache.hadoop.hdfs.DataStreamer.createBlockOutputStream(DataStreamer.java:1789)
        at org.apache.hadoop.hdfs.DataStreamer.nextBlockOutputStream(DataStreamer.java:1743)
        at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:718)
2020-07-31 18:12:37,948 WARN hdfs.DataStreamer: Abandoning BP-1721849101-192.168.0.86-1595473704426:blk_1073774853_34033
2020-07-31 18:12:37,988 WARN hdfs.DataStreamer: Excluding datanode DatanodeInfoWithStorage[192.168.0.174:9866,DS-fa34f00b-2c03-4d0e-ad6e-3a2555735cbd,DISK]
2020-07-31 18:12:38,034 WARN hdfs.DataStreamer: DataStreamer Exception
org.apache.hadoop.ipc.RemoteException(java.io.IOException): File /tmp/input/test.txt._COPYING_ could only be written to 0 of the 1 minReplication nodes. There are 3 da
        at org.apache.hadoop.hdfs.server.blockmanagement.BlockManager.chooseTarget4NewBlock(BlockManager.java:2223)
        at org.apache.hadoop.hdfs.server.namenode.FSDirWriteFileOp.chooseTargetForNewBlock(FSDirWriteFileOp.java:346)
        at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getAdditionalBlock(FSNamesystem.java:2727)
        at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.addBlock(NameNodeRpcServer.java:879)
        at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolServerSideTranslatorPB.addBlock(ClientNamenodeProtocolServerSideTranslatorPB.java:596)
        at org.apache.hadoop.hdfs.protocol.proto.ClientNamenodeProtocolProtos$ClientNamenodeProtocol$2.callBlockingMethod(ClientNamenodeProtocolProtos.java)
        at org.apache.hadoop.ipc.ProtobufRpcEngine$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine.java:530)
        at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:1036)
```

Cause Analysis

The error message "no route to host" is displayed, and the IP address 192.168 is contained in the error message. That is, the internal network route from the client node to the DataNode in the cluster is unreachable. As a result, the file fails to be uploaded.

Procedure

In the client directory of the client node, find the **hdfs-site.xml** file in the HDFS client configuration directory. Add the **dfs.client.use.datanode.hostname** configuration item to the configuration file, and set the value to **true**.

15.9.31 Insufficient Number of Replicas Is Reported During High Concurrent HDFS Writes

Symptom

File writes to HDFS fail occasionally.

The operation log is as follows:

```
105 | INFO  | IPC Server handler 23 on 25000 | IPC Server handler 23 on 25000, call
org.apache.hadoop.hdfs.protocol.ClientProtocol.addBlock from 192.168.1.96:47728 Call#1461167 Retry#0 |
Server.java:2278
java.io.IOException: File /hive/warehouse/000000_0.835bf64f-4103 could only be replicated to 0 nodes
instead of minReplication (=1).  There are 3 datanode(s) running and 3 node(s) are excluded in this
operation.
```

Cause Analysis

- HDFS has a reservation mechanism for file writing: each block to be written is 128 MB no matter whether the file is 10 MB or 1 GB. If a 10 MB file needs to be written, the file occupies 10 MB of the first block and about 118 MB space will be released. If a 1 GB file needs to be written, HDFS writes the file block by block and releases unused space after the file is written.

- If there are a large number of files to be written concurrently, the disk space for reserved write blocks is insufficient. As a result, the file fails to be written.

Solution

Step 1 Log in to the HDFS WebUI and go to the JMX page of the DataNode.

1. On the native HDFS page, choose **Datanodes**.
2. Locate the target DataNode and click the HTTP address to go to the DataNode details page.
3. Change **datanode.html** in **url** to **jmx**.

Step 2 Search for the **XceiverCount** indicator. If the value of this indicator multiplied by the block size exceeds the DataNode disk capacity, the disk space reserved for block write is insufficient.

Step 3 You can use either of the following methods to solve the problem:

Method 1: Reduce the service concurrency.

Method 2: Combine multiple files into one file to reduce the number of files to be written.

----End

15.9.32 HDFS Client Failed to Delete Overlong Directories

Symptom

When a user runs the **hadoop fs -rm -r -f obs://***<obs_path>* command to delete an OBS directory with an overlong path name, the following error message is displayed:

```
2022-02-28 17:12:45,605 INFO internal.RestStorageService: OkHttp cost 19 ms to apply http request
2022-02-28 17:12:45,606 WARN internal.RestStorageService: Request failed, Response code: 400; Request
ID: 0000017F3F9A8545401491602FC8CAD9; Request path: http://wordcount01-fcq.obs.xxx.ulanqab.xxx.com/
user%2Froot%2F.Trash%2FCurrent
%2Ftest1%2F123456789012345678901234567890123456789012345678901234567890123456789012345678901234567
89012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
67890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123
45678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901
23456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789
01234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567
89012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
67890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123
45678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901
23456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789
01234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567
890123456789012345678901234567890123456789012345678901234567890123456
2022-02-28 17:12:45,606 WARN services.AbstractClient: Storage|1|HTTP+XML|getObjectMetadata||||
2022-02-28 17:12:45|2022-02-28 17:12:45|||400|
2022-02-28 17:12:45,607 INFO log.AccessLogger: 2022-02-28 17:12:45 605|
com.obs.services.internal.RestStorageService|executeRequest|560|OkHttp cost 19 ms to apply http request
2022-02-28 17:12:45 606|com.obs.services.internal.RestStorageService|handleThrowable|221|Request failed,
Response code: 400; Request ID: 0000017F3F9A8545401491602FC8CAD9; Request path: http://wordcount01-
fcq.obs.xxx.ulanqab.xxx.com/user%2Froot%2F.Trash%2FCurrent
%2Ftest1%2F123456789012345678901234567890123456789012345678901234567890123456789012345678901234567
```

```
89012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
67890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123
45678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901
23456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789
01234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567
89012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
67890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123
45678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901
23456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789
01234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567
89012345678901234567890123456789012345678901234567890123456
2022-02-28 17:12:45 606|com.obs.services.AbstractClient|doActionWithResult|404|Storage|1|HTTP+XML|
getObjectMetadata||||2022-02-28 17:12:45|2022-02-28 17:12:45|||400|
```

Cause Analysis

When you run the **rm** command to delete some content from the HDFS, the files or directories are not deleted immediately. Instead, they are moved to the **current** directory (**/user/***${Username}***/.Trash/current**) in the recycle bin.

Solution

You can run the **skipTrash** command to skip the HDFS recycle bin and directly delete the data. Set the **dfs.client.skipTrash.enabled=true** configuration item of the HDFS client.

Step 1 Log in to any master node in the cluster as user **root**.

Step 2 Run the following command to edit the **hdfs-site.xml** file used by HDFS:

vim *Client installation directory***/HDFS/hadoop/etc/hadoop/hdfs-site.xml**

Step 3 Add the following content to the **hdfs-site.xml** file:

```
<property>
<name>dfs.client.skipTrash.enabled</name>
<value>true</value>
</property>
```

Step 4 Run the following command to delete the overlong OBS directory:

hadoop fs -rm -r -f –skipTrash obs://*<obs_path>*

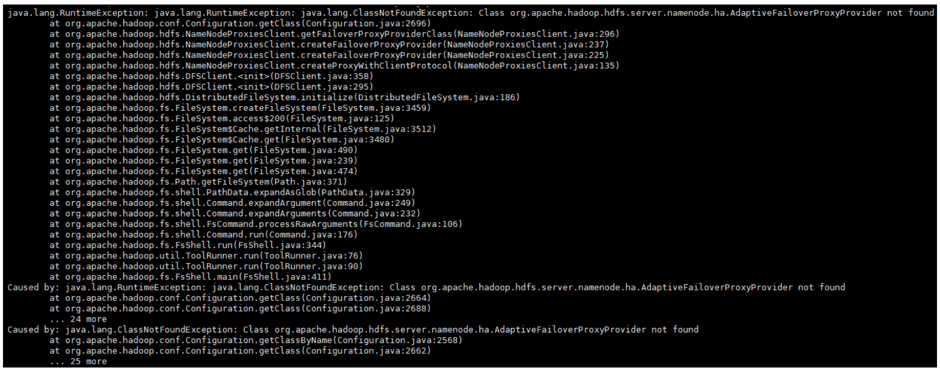
Step 5 Log in to the other master nodes in the cluster and repeat **Step 2** to **Step 4** one by one until the operations are complete on all master nodes.

----End

15.9.33 An Error Is Reported When a Node Outside the Cluster Accesses MRS HDFS

Symptom

When a node outside the MRS cluster accesses HDFS in the cluster, the following error message is reported: Class org.apache.hadoop.hdfs.server.namenode.ha.AdaptiveFailoverProxyProvider not found



```
java.lang.RuntimeException: java.lang.RuntimeException: java.lang.ClassNotFoundException: Class org.apache.hadoop.hdfs.server.namenode.ha.AdaptiveFailoverProxyProvider not found
        at org.apache.hadoop.conf.Configuration.getClass(Configuration.java:2696)
        at org.apache.hadoop.hdfs.NameNodeProxiesClient.getFailoverProxyProviderClass(NameNodeProxiesClient.java:296)
        at org.apache.hadoop.hdfs.NameNodeProxiesClient.createFailoverProxyProvider(NameNodeProxiesClient.java:237)
        at org.apache.hadoop.hdfs.NameNodeProxiesClient.createFailoverProxyProvider(NameNodeProxiesClient.java:225)
        at org.apache.hadoop.hdfs.NameNodeProxiesClient.createProxyWithClientProtocol(NameNodeProxiesClient.java:135)
        at org.apache.hadoop.hdfs.DFSClient.<init>(DFSClient.java:358)
        at org.apache.hadoop.hdfs.DFSClient.<init>(DFSClient.java:295)
        at org.apache.hadoop.hdfs.DistributedFileSystem.initialize(DistributedFileSystem.java:186)
        at org.apache.hadoop.fs.FileSystem.createFileSystem(FileSystem.java:3459)
        at org.apache.hadoop.fs.FileSystem.access$200(FileSystem.java:125)
        at org.apache.hadoop.fs.FileSystem$Cache.getInternal(FileSystem.java:3512)
        at org.apache.hadoop.fs.FileSystem$Cache.get(FileSystem.java:3480)
        at org.apache.hadoop.fs.FileSystem.get(FileSystem.java:490)
        at org.apache.hadoop.fs.FileSystem.get(FileSystem.java:239)
        at org.apache.hadoop.fs.FileSystem.get(FileSystem.java:474)
        at org.apache.hadoop.fs.Path.getFileSystem(Path.java:371)
        at org.apache.hadoop.fs.shell.PathData.expandAsGlob(PathData.java:329)
        at org.apache.hadoop.fs.shell.Command.expandArgument(Command.java:249)
        at org.apache.hadoop.fs.shell.Command.expandArguments(Command.java:232)
        at org.apache.hadoop.fs.shell.FsCommand.processRawArguments(FsCommand.java:106)
        at org.apache.hadoop.fs.shell.Command.run(Command.java:176)
        at org.apache.hadoop.fs.FsShell.run(FsShell.java:344)
        at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)
        at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:90)
        at org.apache.hadoop.fs.FsShell.main(FsShell.java:411)
Caused by: java.lang.RuntimeException: java.lang.ClassNotFoundException: Class org.apache.hadoop.hdfs.server.namenode.ha.AdaptiveFailoverProxyProvider not found
        at org.apache.hadoop.conf.Configuration.getClass(Configuration.java:2664)
        at org.apache.hadoop.conf.Configuration.getClass(Configuration.java:2688)
        ... 24 more
Caused by: java.lang.ClassNotFoundException: Class org.apache.hadoop.hdfs.server.namenode.ha.AdaptiveFailoverProxyProvider not found
        at org.apache.hadoop.conf.Configuration.getClassByName(Configuration.java:2568)
        at org.apache.hadoop.conf.Configuration.getClass(Configuration.java:2662)
        ... 25 more
```

Cause Analysis

The possible causes are as follows:

- An error is reported when an open-source HDFS client accesses HDFS of an MRS cluster.
- An error is reported when the JAR package is used to connect to HDFS of the MRS cluster (including connection to HDFS during task submission).

Procedure

Method 1:

Step 1 Locate the HDFS configuration file **hdfs-site.xml** used by the command or JAR package.

Step 2 Modify the **dfs.client.failover.proxy.provider.hacluster** configuration as follows:

```
<property>
<name>dfs.client.failover.proxy.provider.hacluster</name>
<value>org.apache.hadoop.hdfs.server.namenode.ha.ConfiguredFailoverProxyProvider</value>
</property>
```

NOTE

You can also delete the preceding configuration items.

Step 3 Save the file and access MRS HDFS again.

----End

Method 2:

Step 1 Download the hadoop-plugins matching the MRS cluster version from the Maven repository.

Step 2 Add the downloaded JAR package to the dependency of the command or JAR package.

----End

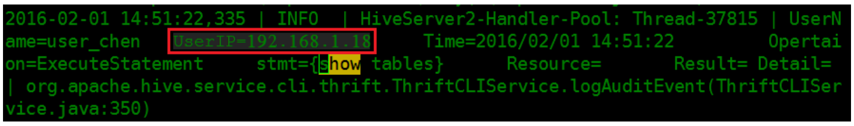
15.10 Using Hive

15.10.1 Content Recorded in Hive Logs

Audit log

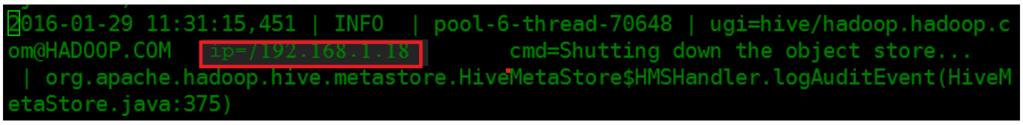
An audit log records at what time a user sends a request to HiveServer and MetaStore from which IP address with what statement.

The following HiveServer audit log shows that at 14:51:22 on February 1, 2016, **user_chen** sent a **show tables** request to HiveServer from the 192.168.1.18 IP address.



```
2016-02-01 14:51:22,335 | INFO  | HiveServer2-Handler-Pool: Thread-37815 | UserN
ame=user_chen   UserIP=192.168.1.18      Time=2016/02/01 14:51:22         Opertai
on=ExecuteStatement     stmt={show tables}      Resource=       Result= Detail=
| org.apache.hive.service.cli.thrift.ThriftCLIService.logAuditEvent(ThriftCLISer
vice.java:350)
```

The following MetaStore audit log shows that user **hive** sent a **shutdown** request to MetaStore from the 192.168.1.18 IP address at 11:31:15 on January 29, 2016.



```
2016-01-29 11:31:15,451 | INFO  | pool-6-thread-70648 | ugi=hive/hadoop.hadoop.c
om@HADOOP.COM   ip=/192.168.1.18        cmd=Shutting down the object store...
 | org.apache.hadoop.hive.metastore.HiveMetaStore$HMSHandler.logAuditEvent(HiveM
etaStore.java:375)
```

Generally, the audit log does not play a role in actual error location. However, the audit log must be checked to solve the following problems:

1. There is no response after a client sends a request. The audit log can be used to check whether the task suspends on the client or server. If the audit log has no related information, the task suspends on the client. If the audit log has related information, view the run log to locate where the program suspends.
2. The audit log can be used to check the number of requests in a specified period of time. You can view the number of requests in a specified period in audit logs.

HiveServer Run Log

HiveServer receives requests from a client (SQL statement), compile and execute the statement (submitted to Yarn or local MapReduce), and interact with MetaStore to obtain metadata information. The HiveServer run log records a complete SQL execution process.

Generally, if SQL statement running fails, check the HiveServer run log first.

MetaStore Run Log

Typically, if the HiveServer run log contains MetaException or MetaStore connection failure, check the MetaStore run log.

GC Log

Both HiveServer and MetaStore have GC logs. If GC-related problems occur, view the GC logs to quickly locate the cause. For example, if HiveServer or MetaStore frequently restarts, check its GC log.

15.10.2 Causes of Hive Startup Failure

The most common cause of the Hive startup failure is that the MetaStore instance cannot connect to DBService. You can view the detailed error information in the MetaStore logs. The reasons for the failure to connect to DBService are as follows:

Possible Cause 1

DBService does not properly initialize the Hive metabase hivemeta.

Procedure 1

Step 1 Run the following commands:

source /opt/Bigdata/MRS_*XXX*/install/dbservice/.dbservice_profile

gsql -h *DBservice floating IP* **-p 20051 -d hivemeta -U hive -W HiveUser@**

Step 2 If the interaction interface cannot be properly displayed, database initialization fails. If the following error information is displayed, the hivemeta configuration may be lost in the configuration file of the node where DBService is located.

```
org.postgresql.util.PSQLException: FATAL: no pg_hba.conf entry for host "192.168.0.146", database
"HIVEMETA"
```

Step 3 Edit **/srv/BigData/dbdata_service/data/pg_hba.conf** by adding **host hivemeta hive 0.0.0.0/0 sha256** to the file.

Step 4 Run the **source /opt/Bigdata/MRS_*XXX*/install/dbservice/.dbservice_profile** command to configure environment variables.

Step 5 Run **gs_ctl -D $GAUSSDATA reload #** to make new configurations take effect.

----End

Possible Cause 2

The floating IP address of DBService is incorrect. As a result, the IP address of the MetaStore node fails to connect to or build mutual trust with the floating IP address, causing MetaStore startup failure.

Procedure 2

The floating IP address of DBService must be an IP address that is not used in the same network segment and cannot be pinged before configuration. Modify the floating IP address of DBService.

15.10.3 "Cannot modify *xxx* at runtime" Is Reported When the set Command Is Executed in a Security Cluster

Symptom

The following error is reported when running the **set** command:

```
0: jdbc:hive2://192.168.1.18:21066/> set mapred.job.queue.name=QueueA;
 Error: Error while processing statement: Cannot modify mapred.job.queue.name at list of params that are
allowed to be modified at runtime (state=42000,code=1)
```

Procedure

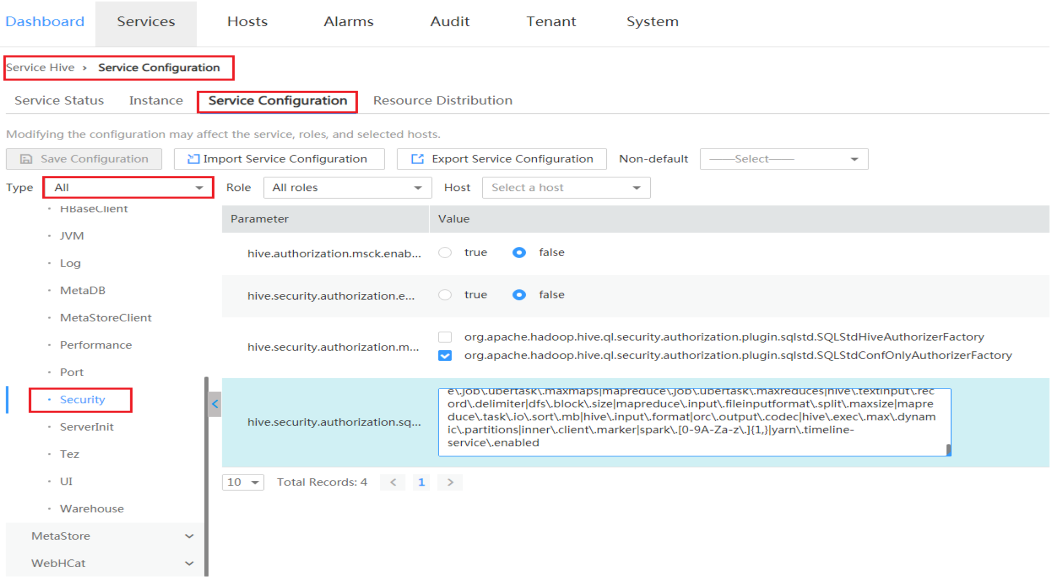
Solution 1:

Step 1 Log in to Manager and modify Hive parameters.

- MRS Manager: Log in to MRS Manager and choose **Services** > **Hive** > **Service Configuration**. Set **Type** to **All** and choose **HiveServer** > **Security**.
- FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Services** > **Hive** > **Configurations** > **All Configurations** > **HiveServer** > **Security**.

Step 2 Add the command parameters to be executed to the **hive.security.authorization.sqlstd.confwhitelist.append** configuration item.

Step 3 Click **Save** and restart **HiveServer**.



Modifying the configuration may affect the service, roles, and selected hosts.

Non-default: ——Select——

Type: All   Role: All roles   Host: Select a host

Parameter / Value

hive.authorization.msck.enab... ○ true ● false

hive.security.authorization.e... ○ true ● false

hive.security.authorization.m... □ org.apache.hadoop.hive.ql.security.authorization.plugin.sqlstd.SQLStdHiveAuthorizerFactory ☑ org.apache.hadoop.hive.ql.security.authorization.plugin.sqlstd.SQLStdConfOnlyAuthorizerFactory

hive.security.authorization.sq...

```
e\.job\.ubertask\.maxmaps|mapreduce\.job\.ubertask\.maxreduces|hive\.textinput\.rec
ord\.delimiter|dfs\.block\.size|mapreduce\.input\.fileinputformat\.split\.maxsize|mapre
duce\.task\.io\.sort\.mb|hive\.input\.format|orc\.output\.codec|hive\.exec\.max\.dynam
ic\.partitions|inner\.client\.marker|spark\.[0-9A-Za-z\.]{1,}|yarn\.timeline-
service\.enabled
```

10   Total Records: 4   < 1 >

----End

Solution 2:

Step 1 Log in to Manager and modify Hive parameters.

- MRS Manager: Log in to MRS Manager and choose **Services** > **Hive** > **Service Configuration**. Set **Type** to **All** and choose **HiveServer** > **Security**.
- FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Services** > **Hive** > **Configurations** > **All Configurations** > **HiveServer** > **Security**.

Step 2 Locate **hive.security.whitelist.switch** and select **OFF**. Click **Save** and restart HiveServer.

----End

15.10.4 How to Specify a Queue When Hive Submits a Job

Symptom

How do I specify a queue when Hive submits a job?

Procedure

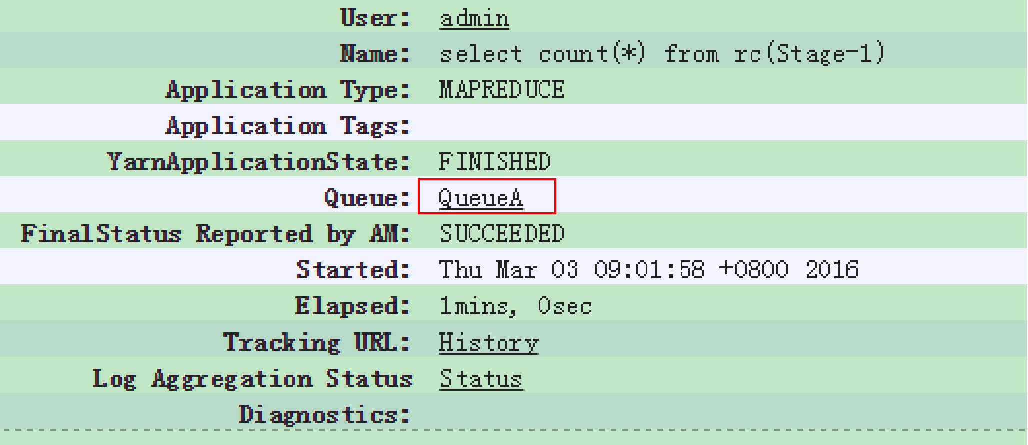
Step 1 Before submitting the job, set the job queue, for example, submitting the job to QueueA.

```
set mapred.job.queue.name=QueueA;
select count(*) from rc;
```

NOTE

The queue name is case sensitive. For example, in this example, **queueA** and **Queuea** are invalid. In addition, the queue must be a leaf queue, and jobs cannot be submitted to a non-leaf queue.

Step 2 After job submission, go to the Yarn page to check the job. The job has been submitted to QueueA.



User:	admin
Name:	select count(*) from rc(Stage-1)
Application Type:	MAPREDUCE
Application Tags:	
YarnApplicationState:	FINISHED
Queue:	QueueA
FinalStatus Reported by AM:	SUCCEEDED
Started:	Thu Mar 03 09:01:58 +0800 2016
Elapsed:	1mins, 0sec
Tracking URL:	History
Log Aggregation Status	Status
Diagnostics:	

----End

15.10.5 How to Set Map and Reduce Memory on the Client

Symptom

How do I set Map and Reduce memory on the client?

Procedure

Before SQL statement execution, run the set command to set parameters of clients related to Map/Reduce.

The following parameters are related to Map and Reduce memory:

```
set mapreduce.map.memory.mb=4096; //Memory required by each Map task
set mapreduce.map.java.opts=-Xmx3276M; //Maximum memory used by the JVM of each Map task
set mapreduce.reduce.memory.mb=4096; //Memory required by each Reduce task
set mapreduce.reduce.java.opts=-Xmx3276M; //Maximum memory used by the JVM of each Reduce task
set mapred.child.java.opts=-Xms1024M -Xmx3584M; // This parameter is a global parameter, which is used
to set Map and Reduce in a unified manner.
```

NOTE

Parameter settings take effect for the current session only.

15.10.6 Specifying the Output File Compression Format When Importing a Table

Question

How do I specify an output file compression format when importing a table?

Procedure

Hive supports the following compression formats:

```
org.apache.hadoop.io.compress.BZip2Codec
org.apache.hadoop.io.compress.Lz4Codec
org.apache.hadoop.io.compress.DeflateCodec
org.apache.hadoop.io.compress.SnappyCodec
org.apache.hadoop.io.compress.GzipCodec
```

- If global settings are required, that is, all tables need to be compressed, you can perform the following global settings for Hive service configuration parameters on the Manager page:
  - Set **hive.exec.compress.output** to **true**.
  - Set **mapreduce.output.fileoutputformat.compress.codec** to **org.apache.hadoop.io.compress.BZip2Codec**.

  NOTE

  The following parameters take effect only when **hive.exec.compress.output** is set to **true**.
- If it needs to be set at the session level, configure the parameters as follows before command execution:

  ```
  set hive.exec.compress.output=true;
  set mapreduce.output.fileoutputformat.compress.codec=org.apache.hadoop.io.compress.SnappyCodec;
  ```

15.10.7 desc Table Cannot Be Completely Displayed

Symptom

How do I make sure that the description is completely displayed when the desc table is too long?

Procedure

Step 1 When starting Beeline of Hive, set **maxWidth** to **20000**.

```
[root@192-168-1-18 logs]# beeline --maxWidth=20000
scan complete in 3ms
Connecting to
...
Beeline version 1.1.0 by Apache Hive
```

Step 2 (Optional) Run the **beeline -help** command to view the client display settings.

```
 -u <database url>              the JDBC URL to connect to
 -n <username>                  the username to connect as
 -p <password>                  the password to connect as
 -d <driver class>              the driver class to use
 -i <init file>                 script file for initialization
 -e <query>                     query that should be executed
 -f <exec file>                 script file that should be executed
```

```
  --hiveconf property=value       Use value for given property
  --color=[true/false]             control whether color is used for display
  --showHeader=[true/false]       show column names in query results
  --headerInterval=ROWS;          the interval between which heades are displayed
  --fastConnect=[true/false]       skip building table/column list for tab-completion
  --autoCommit=[true/false]       enable/disable automatic transaction commit
  --verbose=[true/false]           show verbose error messages and debug info
  --showWarnings=[true/false]     display connection warnings
  --showNestedErrs=[true/false]    display nested errors
  --numberFormat=[pattern]        format numbers using DecimalFormat pattern
  --force=[true/false]             continue running script even after errors
  --maxWidth=MAXWIDTH            the maximum width of the terminal
  --maxColumnWidth=MAXCOLWIDTH    the maximum width to use when displaying columns
  --silent=[true/false]            be more silent
  --autosave=[true/false]          automatically save preferences
  --outputformat=[table/vertical/csv2/tsv2/dsv/csv/tsv]  format mode for result display
                                  Note that csv, and tsv are deprecated - use csv2, tsv2 instead
  --truncateTable=[true/false]    truncate table column when it exceeds length
  --delimiterForDSV=DELIMITER     specify the delimiter for delimiter-separated values output format
(default: |)
  --isolation=LEVEL                set the transaction isolation level
  --nullemptystring=[true/false]  set to true to get historic behavior of printing null as empty string
  --socketTimeOut=n                socket connection timeout interval, in second. The default value is 300.
```

----End

15.10.8 NULL Is Displayed When Data Is Inserted After the Partition Column Is Added

Symptom

1. Run the following command to create a table:

   ```
   create table test_table(
   col1 string,
   col2 string
   )
   PARTITIONED BY(p1 string)
   STORED AS orc tblproperties('orc.compress'='SNAPPY');
   ```

2. Modify the table structure, add partitions, and insert data.

   ```
   alter table test_table add partition(p1='a');
   insert into test_table partition(p1='a') select col1,col2 from temp_table;
   ```

3. Modify the table structure, add columns, and insert data.

   ```
   alter table test_table add columns(col3 string);
   insert into test_table partition(p1='a') select col1,col2,col3 from temp_table;
   ```

4. Query data in the **test_table** table. In the returned result, the values in the **col3** column are all NULL.

   ```
   select * from test_table where p1='a'
   ```

5. Add a table partition and insert data.

   ```
   alter table test_table add partition(p1='b');
   insert into test_table partition(p1='b') select col1,col2,col3 from temp_table;
   ```

6. Query data in the **test_table** table. In the returned result, the value of **col3** is not all NULL.

   ```
   select * from test_table where p1='b'
   ```

Cause Analysis

RESTRICT is the default option for altering a table. In the RESTRICT mode, only the metadata is changed, while the table's partition structure created before the altering operation remains unchanged. However, new partitions created after the altering operation are changed. Therefore, when values of the old partitions are queried, they are all NULL.

Procedure

Add the **cascade** keyword when adding columns, for example:

```
alter table test_table add columns(col3 string) cascade;
```

15.10.9 A Newly Created User Has No Query Permissions

Symptom

When a user is created, an error message is displayed indicating that the user does not have permissions to query data.

```
Error: Error while compiling statement: FAILED: HiveAccessControlException Permission denied: Principal
[name=hive, type=USER] does not have following privileges for operation QUERY [[SELECT] on Object
[type=TABLE_OR_VIEW, name=default.t1]] (state=42000,code=40000)
```

Cause Analysis

The newly created user does not have the permission to operate the Hive component.

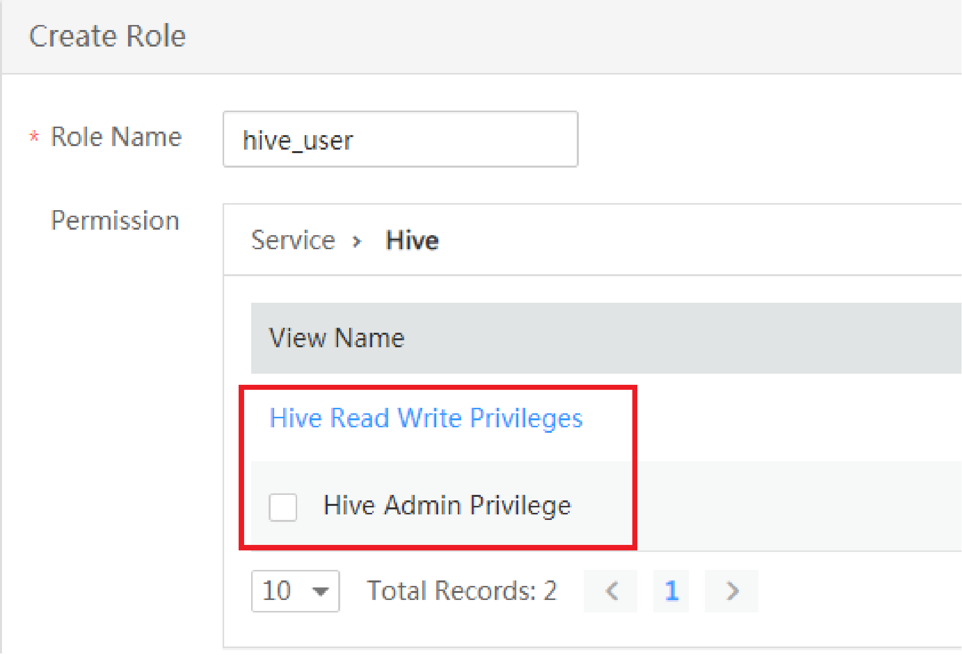
Solution

MRS Manager:

Step 1 Log in to MRS Manager and choose **System** > **Manage Role** > **Create Role**.

Step 2 Enter a role name.

Step 3 In the **Permission** area, select **Hive**. The Hive administrator permission and the read and write permission for Hive tables are displayed.



* Role Name: hive_user

Step 4 Select **Hive Read Write Privileges**. All databases in the Hive column are displayed.

Step 5 Select the permissions required by the role and click **OK**.

Step 6 On MRS Manager, choose **System** > **Manage User**.

Step 7 Locate the row that contains the created user, and click **Modify** in the **Operation** column.

Step 8 Click **Select and Join User Group**. To use the Hive service, you must add a Hive group.

Step 9 Click **Select and Add Role** and select the role created in **Step 5**.

Step 10 Click **OK**.

----End

FusionInsight Manager:

Step 1 Log in to FusionInsight Manager. Choose **System** > **Permission** > **Role**.

Step 2 Click **Create Role**, and set **Role name** and **Description**.

Step 3 Set **Configure Resource Permission** for the role and select **Hive Read and Write Permission** for the Hive table. All databases in the Hive column are displayed.

Step 4 Select the permissions required by the role and click **OK**.

Step 5 On FusionInsight Manager, choose **System** > **Permission** > **User**.

Step 6 Locate the row that contains the created user, and click **Modify** in the **Operation** column.

Step 7 Click **Add** on the right of **User Group**. To use the Hive service, you must add a Hive group.

Step 8 Click **Add** on the right of **Role** and select the role created in 4.

Step 9 Click **OK**.

----End

15.10.10 An Error Is Reported When SQL Is Executed to Submit a Task to a Specified Queue

Symptom

The following error is reported when executing SQL to submit a task to Yarn:

```
Failed to submit application_1475400939788_0033 to YARN :
org.apache.hadoop.security.AccessControlException: User newtest cannot submit applications to queue
root.QueueA
```

Cause Analysis

The current login user does not have the permission to submit the YARN queue.

Solution

Grant the submission permission of the specified Yarn queue to the user. On Manager, choose **System** > **Permission** > **User** and bind a role with the queue submission permission to the user.

15.10.11 An Error Is Reported When the "load data inpath" Command Is Executed

Symptom

The following errors are reported when the **load data inpath** command is executed:

- Error 1:
  ```
  HiveAccessControlException Permission denied. Principal [name=user1, type=USER] does not have
  following privileges on Object [type=DFS_URI, name=hdfs://hacluster/tmp/input/mapdata] for
  operation LOAD : [OBJECT OWNERSHIP]
  ```
- Error 2:
  ```
  HiveAccessControlException Permission denied. Principal [name=user1, type=USER] does not have
  following privileges on Object [type=DFS_URI, name=hdfs://hacluster/tmp/input/mapdata] for
  operation LOAD : [INSERT, DELETE]
  ```
- Error 3:
  ```
  SemanticException [Error 10028]: Line 1:17 Path is not legal ''file:///tmp/input/mapdata'': Move from:
  file:/tmp/input/mapdata to: hdfs://hacluster/user/hive/warehouse/tmp1 is not valid. Please check that
  values for params "default.fs.name" and "hive.metastore.warehouse.dir" do not conflict.
  ```

Cause Analysis

The current login user does not have the permission to operate the directory or the file directory format is incorrect.

Solution

Hive has the following requirements on the **load data inpath** command:

- The file owner must be the user who executes the command.
- The current user must have read and write permissions for the file.
- The current user must have permissions to execute the directory of the file.
- The current user must have the write permission on the directory of the table, because the load operation moves the file to the directory.
- The file format must be the same as the storage format specified by the table. For example, if **stored as rcfile** is specified during table creation but the file format is TXT, it is unsatisfied.
- The file must be stored in HDFS. Files in the local file system cannot be specified using the **file://** form.
- The file name cannot start with an underscore (_) or period (.). A file whose name starts with an underscore (_) or period (.) will be ignored.

  The following shows permissions required when user **test_hive** loads data.

  ```
  [root@192-168-1-18 duan]# hdfs dfs -ls /tmp/input2
  16/03/21 14:45:07 INFO hdfs.PeerCache: SocketCache disabled.
  Found 1 items
  -rw-r--r--   3 test_hive hive          6 2016-03-21 14:44 /tmp/input2/input.txt
  ```

15.10.12 An Error Is Reported When the "load data local inpath" Command Is Executed

Symptom

The following errors are reported when the **load data local inpath** command is executed:

- Error 1:
  ```
  HiveAccessControlException Permission denied. Principal [name=user1, type=USER] does not have following privileges on Object [type=LOCAL_URI, name=file:/tmp/input/mapdata] for operation LOAD : [SELECT, INSERT, DELETE]
  ```
- Error 2:
  ```
  HiveAccessControlException Permission denied. Principal [name=user1, type=USER] does not have following privileges on Object [type=LOCAL_URI, name=file:/tmp/input/mapdata] for operation LOAD : [OBJECT OWNERSHIP]
  ```
- Error 3:
  ```
  SemanticException Line 1:23 Invalid path ''/tmp/input/mapdata'': No files matching path file:/tmp/input/mapdata
  ```

Cause Analysis

The current user does not have the permission to operate the directory or the directory does not exist on the node where HiveServer is located.

Solution

NOTE

Generally, you are not advised to use local files to load data to Hive tables. You are advised to store local files in HDFS and then load data from the cluster.

Hive has the following requirements on the **load data local inpath** command:

- The file must be stored on the HiveServer node, because all commands are sent to the active HiveServer for execution.
- User **omm** must have the read permission for the file and read and execution permissions for the directory where the file is located, because the HiveServer process is started by user **omm** in the OS.
- The file owner must be the user who executes the command.
- The current user must have read and write permissions for the file.
- The file format must be the same as the storage format specified by the table. For example, if **stored as rcfile** is specified during table creation but the file format is TXT, it is unsatisfied.
- The file name cannot start with an underscore (_) or period (.). A file whose name starts with an underscore (_) or period (.) will be ignored.

15.10.13 An Error Is Reported When the "create external table" Command Is Executed

Symptom

The following error is reported when the **create external table** *xx(xx int)* **stored as textfile location '/tmp/aaa/aaa'** command is executed.

```
Permission denied. Principal [name=fantasy, type=USER] does not have following privileges on Object
[type=DFS_URI, name=/tmp/aaa/aaa] for operation CREATETABLE : [SELECT, INSERT, DELETE, OBJECT
OWNERSHIP] (state=42000,code=40000)
```

Cause Analysis

The current login user does not have the read and write permissions for the directory or its parent directory. When an external table is created, whether the current user is checked for its read and write permissions for the specified directory and its subdirectories and subfiles. If the specified directory does not exist, permissions for the parent directory are checked, and so on. If the check results show that the user has no permissions on any directory, "insufficient permission" is reported instead of "The specified directory does not exist".

Solution

Check whether the current user has read and write permissions for the **/tmp/aaa/aaa** path. If the path does not exist, check whether the user has read and write permissions for its parent directory.

15.10.14 An Error Is Reported When the dfs -put Command Is Executed on the Beeline Client

Symptom

Run the following command:

dfs -put /opt/kv1.txt /tmp/kv1.txt

The following error is reported:

```
Permission denied. Principal [name=admin, type=USER] does not have following privileges
onObject[type=COMMAND_PARAMS,name=[-put, /opt/kv1.txt, /tmp/kv1.txt]] for operation DFS : [ADMIN
PRIVILEGE] (state=,code=1)
```

Cause Analysis

The current login user does not have the permissions to run the command.

Solution

If the current user has the **admin** role, run the **set role admin** command to switch to the **admin** role. If the user does not have the admin role, bind the user with the permissions of the corresponding role on the Manager page.

15.10.15 Insufficient Permissions to Execute the set role admin Command

Symptom

When a user runs the following command:

set role admin

The following error is reported:

```
0: jdbc:hive2://192.168.42.26:21066/> set role admin;
Error: Error while processing statement: FAILED: Execution Error, return code 1 from
org.apache.hadoop.hive.ql.exec.DDLTask. dmp_B doesn't belong to role admin (state=08S01,code=1)
```

Cause Analysis

The current user does not have the permissions of the **admin** role of Hive.

Solution

Step 1 Log in to Manager.

- For versions earlier than MRS 3.x, go to **Step 7**.
- For MRS 3.x or later, choose **Cluster** > **Services** > **Hive**. In the upper right corner of the **Dashboard** page, click **More** and check whether **Enable Ranger** is unavailable.
  - If yes, go to **Step 2**.
  - If no, go to **Step 7**.

Step 2 Choose **Cluster** > **Services** > **Ranger** and click **RangerAdmin** in the **Basic Information** area. The Ranger web UI is displayed.

Step 3 Click the username in the upper right corner, select **Log Out** to log out of the system, and log in to the system as user **rangeradmin**.

Step 4 On the homepage, click **Settings** and choose **Roles**.

Step 5 Click the role with **Role Name** set to **admin**. In the **Users** area, click **Select User** and select a username.

Step 6 Click **Add Users**, select **Is Role Admin** in the row where the username is located, and click **Save**.

Step 7 Choose **System** > **Permission** > **Role** and add a role with the Hive administrator permission.

Step 8 On FusionInsight Manager, choose **System** > **Permission** > **User**.

Step 9 In the **Operation** column of the user, click **Modify**.

Step 10 Bind a role that has the Hive administrator permissions to the user and click **OK**.

----End

15.10.16 An Error Is Reported When UDF Is Created Using Beeline

Symptom

Run the following command:

create function fn_test3 as 'test.MyUDF' using jar 'hdfs:///tmp/udf2/MyUDF.jar'

The following error is reported:

```
Error: Error while compiling statement: FAILED: HiveAccessControlException Permission denied: Principal
[name=admin, type=USER] does not have following privileges for operation CREATEFUNCTION [[ADMIN
PRIVILEGE] on Object [type=DATABASE, name=default], [ADMIN PRIVILEGE] on Object [type=FUNCTION,
name=default.fn_test3]] (state=42000,code=40000)
```

Cause Analysis

To create a permanent function in Hive, role **admin** is required.

Solution

Run the **set role admin** command before running the statement.

15.10.17 Difference Between Hive Service Health Status and Hive Instance Health Status

Question

What is the difference between Hive service health status and Hive instance health status?

Solution

The Hive service health status is displayed on the **Services** page and has four values: **Good**, **Bad**, **Partially Healthy**, and **Unknown**. It depends not only on Hive service availability but also the service status of other related components. Simple SQL is used to check Hive service availability.

Hive instances consist of HiveServer and MetaStore. Their health status is determined by communications between instances and JMX and can be **Good** (normal communications), **Concerning** (abnormal communications), or **Unknown** (no communications).

15.10.18 Hive Alarms and Triggering Conditions

Hive Alarms

Alarm ID	Alarm Severity	Auto Clear	Alarm Name	Alarm Type
16000	Minor	TRUE	Percentage of Sessions Connected to the HiveServer to Maximum Number Allowed Exceeds the Threshold	Fault alarm
16001	Minor	TRUE	Hive Warehouse Space Usage Exceeds the Threshold	Fault alarm
16002	Minor	TRUE	The Successful Hive SQL Operations Lower than The Threshold	Fault alarm
16004	Critical	TRUE	Hive Service Unavailable	Fault alarm

Alarm Triggering Scenarios

- 16000: An alarm is triggered when the ratio of the number of sessions connected to HiveServer to the allowed total number of sessions exceeds the threshold. For example, if the number of connected sessions is 9, the allowed total number of sessions is 12, and the threshold is 70%, an alarm is triggered, because 9/12 > 70%.
- 16001: An alarm is triggered when the ratio of HDFS capacities used by Hive to total HDFS capacities allocated to Hive exceeds the threshold. For example, if 500 GB is allocated to Hive, Hive uses 400 GB, and the threshold is 75%, an alarm is triggered, because 400/500 > 75%.
- 16002: An alarm is triggered when SQL execution success rate is lower than the threshold. If two out of four SQL statements are executed successfully and the threshold is 60%, an alarm is triggered, because 2/4 < 60%.
- 16004: An alarm is triggered when the health status of the Hive service changes to Bad.

NOTE

- MRS Manager: To set the alarm threshold, alarm severity, and alarm triggering time segment, choose **System > Configure Alarm Threshold** on MRS Manager.FusionInsight Manager: Choose **O&M** > **Alarm** > **Thresholds** to set the alarm threshold, alarm severity, and alarm triggering time range.
- Metrics related to Hive running can be viewed on the Hive monitoring interface.

15.10.19 "authentication failed" Is Displayed During an Attempt to Connect to the Shell Client

Symptom

In clusters in security mode, the **beeline** command fails to be executed on the Shell client when the HiveServer service is normal, and the system prompts "authentication failed". The following information is displayed.

```
Debug is true storeKey false useTicketCache true useKeyTab false doNotPrompt false ticketCache is null
isInitiator true KeyTab is null refreshKrb5Config is false principal is null tryFirstPass is false useFirstPass is
false storePass is false clearPass is false
Acquire TGT from Cache
Credentials are no longer valid
Principal is null
null credentials from Ticket Cache
[Krb5LoginModule] authentication failed
No password provided
```

Cause Analysis

- The client user does not perform security authentication.
- Kerberos authentication expired.

Solution

Step 1 Log in to the node where the Hive client is installed.

Step 2 Run the **source** *Cluster client installation directory***/bigdata_env** command.

Run the **klist** command to check whether there is a valid ticket in the local end. The following information shows that the ticket became valid at 14:11:42 on December 24, 2016, and expired at 14:11:40 on December 25, 2016. In the period of time, the ticket was available.

```
klist
Ticket cache: FILE:/tmp/krb5cc_0
Default principal: xxx@HADOOP.COM
Valid starting     Expires            Service principal
12/24/16 14:11:42  12/25/16 14:11:40  krbtgt/HADOOP.COM@HADOOP.COM
```

Step 3 Run the **kinit** ***username*** command for authentication and log in to the client again.

----End

15.10.20 Failed to Access ZooKeeper from the Client

Symptom

In clusters in security mode, when the HiveServer service is normal and SQL is executed by using the JDBC interface to connect to HiveServer, "The ZooKeeper client is AuthFailed" is reported.

```
14/05/19 10:52:00 WARN utils.HAClientUtilDummyWatcher: The ZooKeeper client is AuthFailed
 14/05/19 10:52:00 INFO utils.HiveHAClientUtil: Exception thrown while reading data from znode.The
possible reason may be connectionless. This is recoverable. Retrying..
 14/05/19 10:52:16 WARN utils.HAClientUtilDummyWatcher: The ZooKeeper client is AuthFailed
 14/05/19 10:52:32 WARN utils.HAClientUtilDummyWatcher: The ZooKeeper client is AuthFailed
 14/05/19 10:52:32 ERROR st.BasicTestCase: Exception: Could not establish connection to active hiveserver
 java.sql.SQLException: Could not establish connection to active hiveserver
```

Or an error is reported stating "Unable to read HiveServer2 configs from ZooKeeper":

```
Exception in thread "main" java.sql.SQLException: org.apache.hive.jdbc.ZooKeeperHiveClientException:
Unable to read HiveServer2 configs from ZooKeeper
at org.apache.hive.jdbc.HiveConnection.<init>(HiveConnection.java:144)
at org.apache.hive.jdbc.HiveDriver.connect(HiveDriver.java:105)
at java.sql.DriverManager.getConnection(DriverManager.java:664)
at java.sql.DriverManager.getConnection(DriverManager.java:247)
at JDBCExample.main(JDBCExample.java:82)
Caused by: org.apache.hive.jdbc.ZooKeeperHiveClientException: Unable to read HiveServer2 configs from
ZooKeeper
at org.apache.hive.jdbc.ZooKeeperHiveClientHelper.configureConnParams(ZooKeeperHiveClientHelper.java:
100)
at org.apache.hive.jdbc.Utils.configureConnParams(Utils.java:509)
at org.apache.hive.jdbc.Utils.parseURL(Utils.java:429)
at org.apache.hive.jdbc.HiveConnection.<init>(HiveConnection.java:142)
... 4 more
Caused by: org.apache.zookeeper.KeeperException$ConnectionLossException: KeeperErrorCode =
ConnectionLoss for /hiveserver2
at org.apache.zookeeper.KeeperException.create(KeeperException.java:99)
at org.apache.zookeeper.KeeperException.create(KeeperException.java:51)
at org.apache.zookeeper.ZooKeeper.getChildren(ZooKeeper.java:2374)
at org.apache.curator.framework.imps.GetChildrenBuilderImpl$3.call(GetChildrenBuilderImpl.java:214)
at org.apache.curator.framework.imps.GetChildrenBuilderImpl$3.call(GetChildrenBuilderImpl.java:203)
at org.apache.curator.RetryLo, op.callWithRetry(RetryLoop.java:107)
at
org.apache.curator.framework.imps.GetChildrenBuilderImpl.pathInForeground(GetChildrenBuilderImpl.java:
200)
at org.apache.curator.framework.imps.GetChildrenBuilderImpl.forPath(GetChildrenBuilderImpl.java:191)
at org.apache.curator.framework.imps.GetChildrenBuilderImpl.forPath(GetChildrenBuilderImpl.java:38)
```

Cause Analysis

- When the client connects to HiveServer, the HiveServer address is automatically obtained from ZooKeeper. If ZooKeeper connection authentication is abnormal, the HiveServer address cannot be obtained from ZooKeeper correctly.
- During ZooKeeper connection authentication, **krb5.conf**, **principal**, **keytab**, and related information must be loaded to the client. Authentication failure causes are as follows:
  - The **user.keytab** path is incorrectly entered.
  - **user.principal** is incorrectly entered.
  - The cluster has switched the domain name. However, the old principal is used when the client combines the URL.

– The client cannot pass Kerberos authentication due to firewall settings. Ports 21730 (TCP), 21731 (TCP/UDP), and 21732 (TCP/UDP) need to be opened for Kerberos.

Solution

Step 1 Ensure that the user can properly access the **user.keytab** file in related paths on the client node.

Step 2 Ensure that the user's **user.principal** corresponds to the specified **keytab** file.

Run the **klist -kt keytabpath/user.keytab** command to check the file.

Step 3 If the cluster has switched the domain name, the **principal** field used in the URL must be the new domain name.

For example, the default value is **hive/hadoop.hadoop.com@HADOOP.COM**. If the cluster has switched the domain name, the field must be changed accordingly. For example, if the domain name is **abc.com**, enter **hive/hadoop.abc.com@ABC.COM**.

Step 4 Ensure that authentication is normal and HiveServer can be connected.

Run the following commands on the client:

source *Client installation directory***/bigdata_env**

kinit *username*

Run the **beeline** command on the client to ensure normal running.

----End

15.10.21 "Invalid function" Is Displayed When a UDF Is Used

Symptom

When a UDF is created on the Hive client using Spark, "Error 10011" indicating "invalid function" is reported:

```
Error: Error while compiling statement: FAILED: SemanticException [Error 10011]: Line 1:7 Invalid function
'test_udf' (state=42000,code=10011)
```

The preceding problem occurs when multiple HiveServers use a UDF. For example, if metadata is not synchronized in time when the UDF created on HiveServer2 is used on HiveServer1, the preceding error is reported when clients on HiveServer1 are connected.

Cause Analysis

Metadata shared by multiple HiveServers or Hive and Spark is not synchronized, causing memory data inconsistency between different HiveServer instances and invalid UDF.

Solution

Synchronize new UDF information to HiveServer and reload the function.

15.10.22 Hive Service Status Is Unknown

Cause Analysis

The Hive service stops.

Solution

Restart the Hive service.

15.10.23 Health Status of a HiveServer or MetaStore Instance Is Unknown

Symptom

The health status of a HiveServer or MetaStore instance is unknown.

Cause Analysis

The HiveServer or MetaStore instance is stopped.

Solution

Restart the HiveServer or MetaStore instance.

15.10.24 Health Status of a HiveServer or MetaStore Instance Is Concerning

Symptom

The health status of the HiveServer or MetaStore instance is **Concerning**.

Cause Analysis

The HiveServer or MetaStore instance cannot be normally started. For example, when modifying the MetaStore/HiveServer GC parameter, you can view the startup log of the corresponding process, for example, the **hiveserver.out(hadoop-omm-jar-192-168-1-18.out)** file. The following exception occurs:

Error: Could not find or load main class Xmx2048M

The preceding information indicates that **Xmx2048M** is used as the startup parameter of the Java process instead of the JVM during the startup of the Java virtual machine. As shown in the following information, the hyphen (-) is deleted mistakenly.

```
METASTORE_GC_OPTS=Xms1024M Xmx2048M -DIgnoreReplayReqDetect
-XX\:CMSFullGCsBeforeCompaction\=1 -XX\:+UseConcMarkSweepGC
-XX\:+CMSParallelRemarkEnabled -XX\:+UseCMSCompactAtFullCollection
-XX\:+ExplicitGCInvokesConcurrent -server -XX\:MetaspaceSize\=128M
-XX\:MaxMetaspaceSize\=256M
```

Solution

Check the latest changes to detect incorrect settings.

```
METASTORE_GC_OPTS=Xms1024M -Xmx2048M -DIgnoreReplayReqDetect
-XX\:CMSFullGCsBeforeCompaction\=1 -XX\:+UseConcMarkSweepGC
-XX\:+CMSParallelRemarkEnabled -XX\:+UseCMSCompactAtFullCollection
-XX\:+ExplicitGCInvokesConcurrent -server -XX\:MetaspaceSize\=128M
-XX\:MaxMetaspaceSize\=256M
```

15.10.25 Garbled Characters Returned upon a select Query If Text Files Are Compressed Using ARC4

Symptom

If a Hive query result table is compressed and stored using the ARC4 algorithm, garbled characters are returned after the select * query is conducted in the result table.

Cause Analysis

The default Hive compression format is not ARC4 or output compression is disabled.

Solution

Step 1 If garbled characters are returned after the SETECT query, set the following in Beeline:

set mapreduce.output.fileoutputformat.compress.codec=org.apache.hadoop.io.encryption.arc4.ARC4BlockCodec;

set hive.exec.compress.output=true;

Step 2 Import the table to a new table using block decompression.

insert overwrite table tbl_result select * from tbl_source;

Step 3 Perform the query again.

select * from tbl_result;

----End

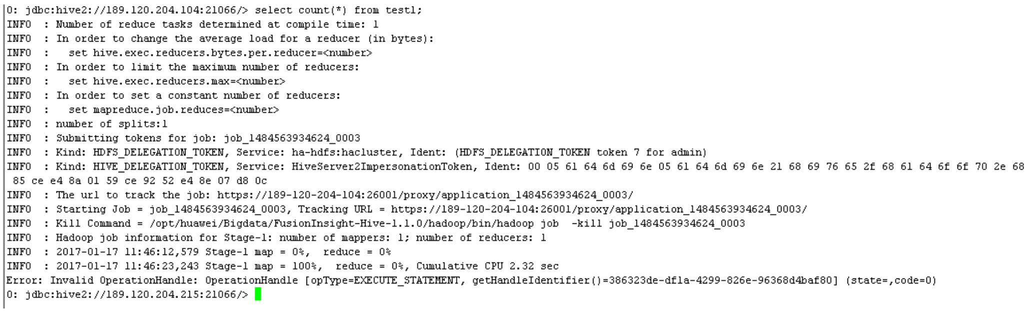
15.10.26 Hive Task Failed to Run on the Client But Successful on Yarn

Symptom

When Hive task running fails, an error similar to the following is reported on the client:

```
Error:Invalid OperationHandler:OperationHander
[opType=EXECUTE_STATEMENT,getHandleIdentifier()=XXX](state=,code=0)
```

However, the MapReduce task that is submitted by the task to Yarn is successfully executed.



```
0: jdbc:hive2://189.120.204.104:21066/> select count(*) from testl;
INFO  : Number of reduce tasks determined at compile time: 1
INFO  : In order to change the average load for a reducer (in bytes):
INFO  :   set hive.exec.reducers.bytes.per.reducer=<number>
INFO  : In order to limit the maximum number of reducers:
INFO  :   set hive.exec.reducers.max=<number>
INFO  : In order to set a constant number of reducers:
INFO  :   set mapreduce.job.reduces=<number>
INFO  : number of splits:1
INFO  : Submitting tokens for job: job_1484563934624_0003
INFO  : Kind: HDFS_DELEGATION_TOKEN, Service: ha-hdfs:hacluster, Ident: (HDFS_DELEGATION_TOKEN token 7 for admin)
INFO  : Kind: HIVE_DELEGATION_TOKEN, Service: HiveServer2ImpersonationToken, Ident: 00 05 61 64 6d 69 6e 05 61 64 6d 69 6e 21 68 69 76 65 2f 68 61 64 6f 6f 70 2e 68
 85 ce e4 8a 01 59 ce 92 52 e4 8e 07 d8 0c
INFO  : The url to track the job: https://189-120-204-104:26001/proxy/application_1484563934624_0003/
INFO  : Starting Job = job_1484563934624_0003, Tracking URL = https://189-120-204-104:26001/proxy/application_1484563934624_0003/
INFO  : Kill Command = /opt/huawei/Bigdata/FusionInsight-Hive-1.1.0/hadoop/bin/hadoop job  -kill job_1484563934624_0003
INFO  : Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
INFO  : 2017-01-17 11:46:12,579 Stage-1 map = 0%,  reduce = 0%
INFO  : 2017-01-17 11:46:23,243 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.32 sec
Error: Invalid OperationHandle: OperationHandle [opType=EXECUTE_STATEMENT, getHandleIdentifier()=386323de-df1a-4299-826e-96368d4baf80] (state=,code=0)
0: jdbc:hive2://189.120.204.215:21066/>
```

Cause Analysis

The cluster where the error occurs has two HiveServer instances. The error in the log of one HiveServer instance is the same as the error (Error: Invalid OperationHandler) reported on the client. In the log of the other HiveServer instance, **START_UP** information similar to the following is printed when the error occurs, which indicates that the process is killed and restarted during that time. Because the HiveServer instance the task process plans to connect to is killed, it connects to the other healthy one, causing the error.

```
2017-02-15 14:40:11,309 | INFO  | main | STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting HiveServer2
STARTUP_MSG:   host = XXX-120-85-154/XXX.120.85.154
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 1.3.0
```

Solution

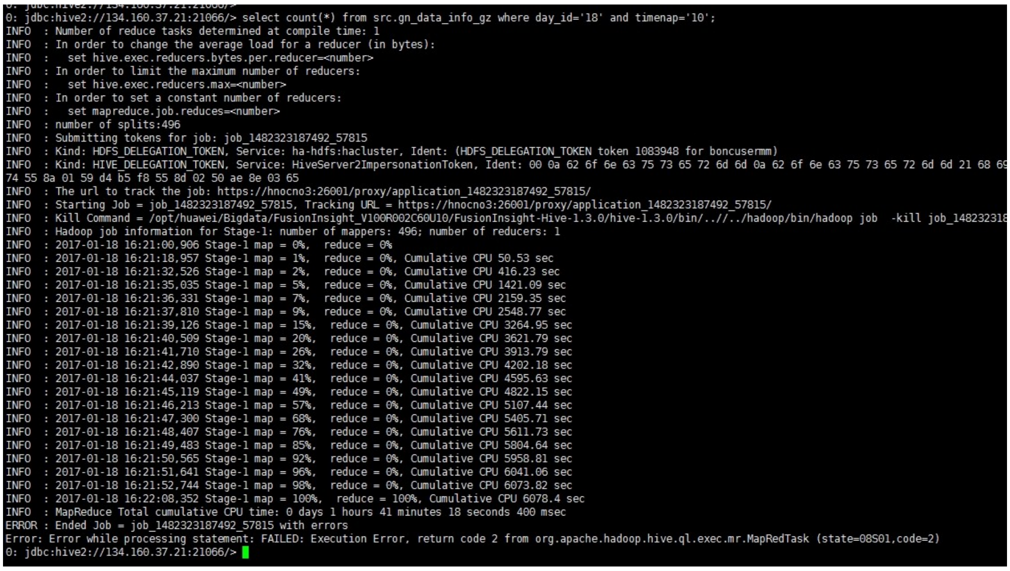
Submit the task again and ensure that the HiveServer process is not manually restarted during task execution.

15.10.27 An Error Is Reported When the select Statement Is Executed

Symptom

When the **select count(*) from XXX** statement is executed, the client reports the error "Error:Error while processing statement :FAILED:Execution Error,return code 2 from...".

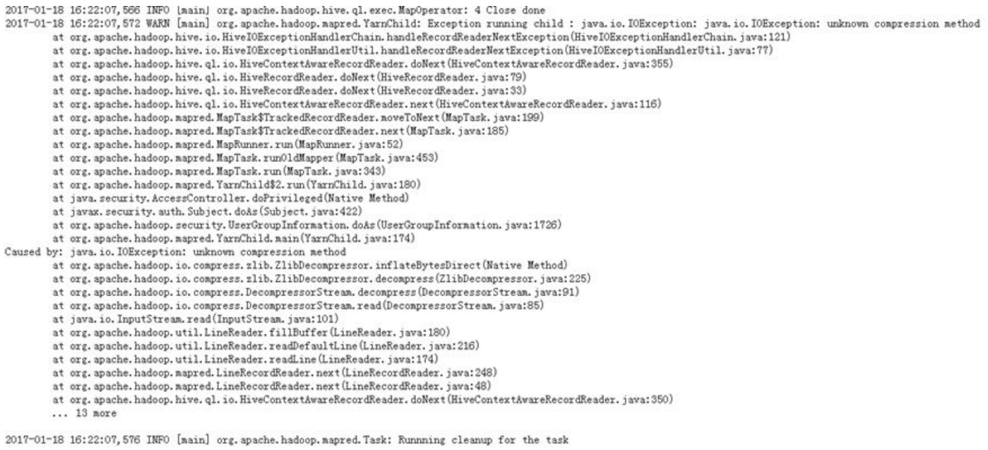
return code 2 indicates that the task fails because an error is reported during the execution of the MapReduce task.



```
0: jdbc:hive2://134.160.37.21:21066/>
0: jdbc:hive2://134.160.37.21:21066/> select count(*) from src.gn_data_info_gz where day_id='18' and timenap='10';
INFO  : Number of reduce tasks determined at compile time: 1
INFO  : In order to change the average load for a reducer (in bytes):
INFO  :   set hive.exec.reducers.bytes.per.reducer=<number>
INFO  : In order to limit the maximum number of reducers:
INFO  :   set hive.exec.reducers.max=<number>
INFO  : In order to set a constant number of reducers:
INFO  :   set mapreduce.job.reduces=<number>
INFO  : number of splits:496
INFO  : Submitting tokens for job: job_1482323187492_57815
INFO  : Kind: HDFS_DELEGATION_TOKEN, Service: ha-hdfs:hacluster, Ident: (HDFS_DELEGATION_TOKEN token 1083948 for boncusermm)
INFO  : Kind: HIVE_DELEGATION_TOKEN, Service: HiveServer2ImpersonationToken, Ident: 00 0a 62 6f 6e 63 75 73 65 72 6d 6d 0a 62 6f 6e 63 75 73 65 72 6d 6d 21 68 69
74 55 8a 01 59 d4 b5 f8 55 8d 02 50 ae 8e 03 65
INFO  : The url to track the job: https://hnocno3:26001/proxy/application_1482323187492_57815/
INFO  : Starting Job = job_1482323187492_57815, Tracking URL = https://hnocno3:26001/proxy/application_1482323187492_57815/
INFO  : Kill Command = /opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Hive-1.3.0/hive-1.3.0/bin/..//../hadoop/bin/hadoop job  -kill job_14823231
INFO  : Hadoop job information for Stage-1: number of mappers: 496; number of reducers: 1
INFO  : 2017-01-18 16:21:00,906 Stage-1 map = 0%,  reduce = 0%
INFO  : 2017-01-18 16:21:18,957 Stage-1 map = 1%,  reduce = 0%, Cumulative CPU 50.53 sec
INFO  : 2017-01-18 16:21:32,526 Stage-1 map = 2%,  reduce = 0%, Cumulative CPU 416.23 sec
INFO  : 2017-01-18 16:21:35,035 Stage-1 map = 5%,  reduce = 0%, Cumulative CPU 1421.09 sec
INFO  : 2017-01-18 16:21:36,331 Stage-1 map = 7%,  reduce = 0%, Cumulative CPU 2159.35 sec
INFO  : 2017-01-18 16:21:37,810 Stage-1 map = 9%,  reduce = 0%, Cumulative CPU 2548.77 sec
INFO  : 2017-01-18 16:21:39,126 Stage-1 map = 15%,  reduce = 0%, Cumulative CPU 3264.95 sec
INFO  : 2017-01-18 16:21:40,509 Stage-1 map = 20%,  reduce = 0%, Cumulative CPU 3621.79 sec
INFO  : 2017-01-18 16:21:41,710 Stage-1 map = 26%,  reduce = 0%, Cumulative CPU 3913.79 sec
INFO  : 2017-01-18 16:21:42,890 Stage-1 map = 32%,  reduce = 0%, Cumulative CPU 4202.18 sec
INFO  : 2017-01-18 16:21:44,037 Stage-1 map = 41%,  reduce = 0%, Cumulative CPU 4595.63 sec
INFO  : 2017-01-18 16:21:45,119 Stage-1 map = 49%,  reduce = 0%, Cumulative CPU 4822.15 sec
INFO  : 2017-01-18 16:21:46,213 Stage-1 map = 57%,  reduce = 0%, Cumulative CPU 5107.44 sec
INFO  : 2017-01-18 16:21:47,300 Stage-1 map = 68%,  reduce = 0%, Cumulative CPU 5405.71 sec
INFO  : 2017-01-18 16:21:48,407 Stage-1 map = 76%,  reduce = 0%, Cumulative CPU 5611.73 sec
INFO  : 2017-01-18 16:21:49,483 Stage-1 map = 85%,  reduce = 0%, Cumulative CPU 5804.64 sec
INFO  : 2017-01-18 16:21:50,565 Stage-1 map = 92%,  reduce = 0%, Cumulative CPU 5958.81 sec
INFO  : 2017-01-18 16:21:51,641 Stage-1 map = 96%,  reduce = 0%, Cumulative CPU 6041.06 sec
INFO  : 2017-01-18 16:21:52,744 Stage-1 map = 98%,  reduce = 0%, Cumulative CPU 6073.82 sec
INFO  : 2017-01-18 16:22:08,352 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6078.4 sec
INFO  : MapReduce Total cumulative CPU time: 0 days 1 hours 41 minutes 18 seconds 400 msec
ERROR : Ended Job = job_1482323187492_57815 with errors
Error: Error while processing statement: FAILED: Execution Error, return code 2 from org.apache.hadoop.hive.ql.exec.mr.MapRedTask (state=08S01,code=2)
0: jdbc:hive2://134.160.37.21:21066/>
```

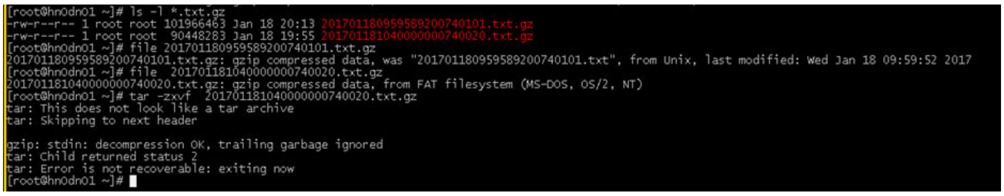
Cause Analysis

1. Go to the native Yarn page to check the MapReduce task logs. The check result shows that the error occurs due to unidentified compression mode. The file name suffix is **.gzip** but the stack reports **.zlib**.



```
2017-01-18 16:22:07,566 INFO [main] org.apache.hadoop.hive.ql.exec.MapOperator: 4 Close done
2017-01-18 16:22:07,572 WARN [main] org.apache.hadoop.mapred.YarnChild: Exception running child : java.io.IOException: java.io.IOException: unknown compression method
        at org.apache.hadoop.hive.io.HiveIOExceptionHandlerChain.handleRecordReaderNextException(HiveIOExceptionHandlerChain.java:121)
        at org.apache.hadoop.hive.io.HiveIOExceptionHandlerUtil.handleRecordReaderNextException(HiveIOExceptionHandlerUtil.java:77)
        at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.doNext(HiveContextAwareRecordReader.java:355)
        at org.apache.hadoop.hive.ql.io.HiveRecordReader.doNext(HiveRecordReader.java:79)
        at org.apache.hadoop.hive.ql.io.HiveRecordReader.doNext(HiveRecordReader.java:33)
        at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.next(HiveContextAwareRecordReader.java:116)
        at org.apache.hadoop.mapred.MapTask$TrackedRecordReader.moveToNext(MapTask.java:199)
        at org.apache.hadoop.mapred.MapTask$TrackedRecordReader.next(MapTask.java:185)
        at org.apache.hadoop.mapred.MapRunner.run(MapRunner.java:52)
        at org.apache.hadoop.mapred.MapTask.runOldMapper(MapTask.java:453)
        at org.apache.hadoop.mapred.MapTask.run(MapTask.java:343)
        at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:180)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1726)
        at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:174)
Caused by: java.io.IOException: unknown compression method
        at org.apache.hadoop.io.compress.zlib.ZlibDecompressor.inflateBytesDirect(Native Method)
        at org.apache.hadoop.io.compress.zlib.ZlibDecompressor.decompress(ZlibDecompressor.java:225)
        at org.apache.hadoop.io.compress.DecompressorStream.decompress(DecompressorStream.java:91)
        at org.apache.hadoop.io.compress.DecompressorStream.read(DecompressorStream.java:85)
        at java.io.InputStream.read(InputStream.java:101)
        at org.apache.hadoop.util.LineReader.fillBuffer(LineReader.java:180)
        at org.apache.hadoop.util.LineReader.readDefaultLine(LineReader.java:216)
        at org.apache.hadoop.util.LineReader.readLine(LineReader.java:174)
        at org.apache.hadoop.mapred.LineRecordReader.next(LineRecordReader.java:248)
        at org.apache.hadoop.mapred.LineRecordReader.next(LineRecordReader.java:48)
        at org.apache.hadoop.hive.ql.io.HiveContextAwareRecordReader.doNext(HiveContextAwareRecordReader.java:350)
        ... 13 more

2017-01-18 16:22:07,576 INFO [main] org.apache.hadoop.mapred.Task: Runnning cleanup for the task
```

2. Therefore, the HDFS file corresponding to the table that is queried may be incorrect. According to the file name printed in the map log, download the file from HDFS to the local end. The file whose name is suffixed with **.gz** fails to be decompressed by running the **tar** command because its format is incorrect. Run the **file** command to check the file property. The command output shows that the file is compressed from the FAT system instead of UNIX.



```
[root@hn0dn01 ~]# ls -l *.txt.gz
-rw-r--r-- 1 root root 101966463 Jan 18 20:13 20170118095958920074O101.txt.gz
-rw-r--r-- 1 root root  90448283 Jan 18 19:55 20170118104000000740020.txt.gz
[root@hn0dn01 ~]# file 20170118095958920074O101.txt.gz
20170118095958920074O101.txt.gz: gzip compressed data, was "20170118095958920074O101.txt", from Unix, last modified: Wed Jan 18 09:59:52 2017
[root@hn0dn01 ~]# file  20170118104000000740020.txt.gz
20170118104000000740020.txt.gz: gzip compressed data, from FAT filesystem (MS-DOS, OS/2, NT)
[root@hn0dn01 ~]# tar -zxvf  20170118104000000740020.txt.gz
tar: This does not look like a tar archive
tar: Skipping to next header

gzip: stdin: decompression OK, trailing garbage ignored
tar: Child returned status 2
tar: Error is not recoverable: exiting now
[root@hn0dn01 ~]#
```

Solution

Delete the file with an incorrect format from the HDFS directory or replace it with a correct one.

15.10.28 Failed to Drop a Large Number of Partitions

Symptom

When the **drop partition** operation is performed, the following information is displayed:

```
MetaStoreClient lost connection. Attempting to reconnect. |
org.apache.hadoop.hive.metastore.RetryingMetaStoreClient.invoke(RetryingMetaStoreClient.java:187)
org.apache.thrift.transport.TTransportException
at org.apache.thrift.transport.TIOStreamTransport.read(TIOStreamTransport.java:132)
at org.apache.thrift.transport.TTransport.xxx(TTransport.java:86)
at org.apache.thrift.transport.TSaslTransport.readLength(TSaslTransport.java:376)
at org.apache.thrift.transport.TSaslTransport.readFrame(TSaslTransport.java:453)
at org.apache.thrift.transport.TSaslTransport.read(TSaslTransport.java:435)
...
```

As indicated by the MetaStore log, StackOverFlow occurs.

```
2017-04-22 01:00:58,834 | ERROR | pool-6-thread-208 | java.lang.StackOverflowError
at org.datanucleus.store.rdbms.sql.SQLText.toSQL(SQLText.java:330)
at org.datanucleus.store.rdbms.sql.SQLText.toSQL(SQLText.java:339)
at org.datanucleus.store.rdbms.sql.SQLText.toSQL(SQLText.java:339)
at org.datanucleus.store.rdbms.sql.SQLText.toSQL(SQLText.java:339)
at org.datanucleus.store.rdbms.sql.SQLText.toSQL(SQLText.java:339)
```

Cause Analysis

The processing logic of the drop partition operation is to find all the partitions that meet the conditions, combine them, and delete them together. However, because the number of partitions is too large and the data stack for deleting metadata is deep, StackOverFlow errors occur.

Solution

Delete partitions in batches.

15.10.29 Failed to Start a Local Task

Symptom

1. When operations such as JOIN are performed for a small amount of data, a local task will be started. However, the execution fails and reports the following error:

```
jdbc:hive2://10.*.*.*:21066/> select a.name ,b.sex from student a join student1 b on (a.name = b.name);
ERROR : Execution failed with exit status: 1
ERROR : Obtaining error information
ERROR :
Task failed!
Task ID:
  Stage-4
...
Error: Error while processing statement: FAILED: Execution Error, return code 1 from
org.apache.hadoop.hive.ql.exec.mr.MapredLocalTask (state=08S01,code=1)
...
```

2. The HiveServer log shows that the local task fails to start.

```
2018-04-25 16:37:19,296 | ERROR | HiveServer2-Background-Pool: Thread-79 | Execution failed with
exit status: 1 | org.apache.hadoop.hive.ql.session.SessionState$LogHelper.printError(SessionState.java:
1016)
2018-04-25 16:37:19,296 | ERROR | HiveServer2-Background-Pool: Thread-79 | Obtaining error
information | org.apache.hadoop.hive.ql.session.SessionState$LogHelper.printError(SessionState.java:
1016)
2018-04-25 16:37:19,297 | ERROR | HiveServer2-Background-Pool: Thread-79 |
Task failed!
Task ID:
  Stage-4
Logs:
 | org.apache.hadoop.hive.ql.session.SessionState$LogHelper.printError(SessionState.java:1016)
2018-04-25 16:37:19,297 | ERROR | HiveServer2-Background-Pool: Thread-79 | /var/log/Bigdata/hive/
hiveserver/hive.log | org.apache.hadoop.hive.ql.session.SessionState
$LogHelper.printError(SessionState.java:1016)
2018-04-25 16:37:19,297 | ERROR | HiveServer2-Background-Pool: Thread-79 | Execution failed with
exit status: 1 |
org.apache.hadoop.hive.ql.exec.mr.MapredLocalTask.executeInChildVM(MapredLocalTask.java:342)
2018-04-25 16:37:19,309 | ERROR | HiveServer2-Background-Pool: Thread-79 | FAILED: Execution
Error, return code 1 from org.apache.hadoop.hive.ql.exec.mr.MapredLocalTask |
org.apache.hadoop.hive.ql.session.SessionState$LogHelper.printError(SessionState.java:1016)
...
2018-04-25 16:37:36,438 | ERROR | HiveServer2-Background-Pool: Thread-88 | Error running hive
query:  | org.apache.hive.service.cli.operation.SQLOperation$1$1.run(SQLOperation.java:248)
org.apache.hive.service.cli.HiveSQLException: Error while processing statement: FAILED: Execution
Error, return code 1 from org.apache.hadoop.hive.ql.exec.mr.MapredLocalTask
        at org.apache.hive.service.cli.operation.Operation.toSQLException(Operation.java:339)
        at org.apache.hive.service.cli.operation.SQLOperation.runQuery(SQLOperation.java:169)
        at org.apache.hive.service.cli.operation.SQLOperation.access$200(SQLOperation.java:75)
        at org.apache.hive.service.cli.operation.SQLOperation$1$1.run(SQLOperation.java:245)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1710)
        at org.apache.hive.service.cli.operation.SQLOperation$1.run(SQLOperation.java:258)
        at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:511)
        at java.util.concurrent.FutureTask.run(FutureTask.java:266)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1142)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:617)
        at java.lang.Thread.run(Thread.java:745)
```

3. The **hs_err_pid_*****.log** file in the HiveServer log directory **/var/log/Bigdata/hive/hiveserver** contains an error about insufficient memory.

```
# There is insufficient memory for the Java Runtime Environment to continue.
# Native memory allocation (mmap) failed to map 20776943616 bytes for committing reserved
memory.
  ...
```

Cause Analysis

When Hive executes JOIN for a small amount of data, MapJoin is generated. During MapJoin execution, a local task is started. JVM memory launched by the local task inherits the memory of the parent process.

When multiple JOIN operations are executed, multiple local tasks are started. If the host is out of memory, the local tasks fail to start.

Solution

Step 1 Go to the Hive configuration page.

- For versions earlier than MRS 2.0.1: Log in to MRS Manager, choose **Services** > **Hive** > **Service Configuration**, and select **All** from the **Basic** drop-down list.

- For MRS 2.0.1 or later: Click the cluster name on the MRS console, choose **Components** > **Hive** > **Service Configuration**, and select **All** from the **Basic** drop-down list

  NOTE

  If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

- For MRS 3.*x* or later: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the target cluster, and choose **Services** > **Hive** > **Configurations** > **All Configurations**.

Step 2 Search for the **hive.auto.convert.join** parameter and change the value of **hive.auto.convert.join** in Hive to **false**. Save the configuration and restart the service.

The value change may deteriorate service performance. You can perform the next step to avoid adverse impacts on the performance.

Step 3 Search for the **HIVE_GC_OPTS** parameter and decrease the value of **Xms** based on service requirements. The minimum value is half that of **Xmx**. After the modification, save the configuration and restart the service.

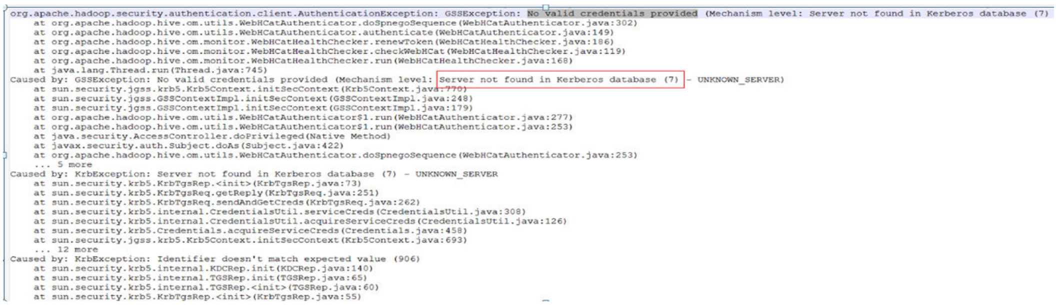
----End

15.10.30 Failed to Start WebHCat

Symptom

WebHCat fails to be started after the hostname is changed.

The following error is reported in the WebHCat startup log (**/var/log/Bigdata/hive/webhcat/hive.log**) of the corresponding node:



```
org.apache.hadoop.security.authentication.client.AuthenticationException: GSSException: No valid credentials provided (Mechanism level: Server not found in Kerberos database (7)
    at org.apache.hadoop.hive.om.utils.WebHCatAuthenticator.doSpnegoSequence(WebHCatAuthenticator.java:302)
    at org.apache.hadoop.hive.om.utils.WebHCatAuthenticator.authenticate(WebHCatAuthenticator.java:149)
    at org.apache.hadoop.hive.om.monitor.WebHCatHealthChecker.renewToken(WebHCatHealthChecker.java:186)
    at org.apache.hadoop.hive.om.monitor.WebHCatHealthChecker.checkWebHCat(WebHCatHealthChecker.java:119)
    at org.apache.hadoop.hive.om.monitor.WebHCatHealthChecker.run(WebHCatHealthChecker.java:168)
    at java.lang.Thread.run(Thread.java:745)
Caused by: GSSException: No valid credentials provided (Mechanism level: Server not found in Kerberos database (7) - UNKNOWN_SERVER)
    at sun.security.jgss.krb5.Krb5Context.initSecContext(Krb5Context.java:770)
    at sun.security.jgss.GSSContextImpl.initSecContext(GSSContextImpl.java:248)
    at sun.security.jgss.GSSContextImpl.initSecContext(GSSContextImpl.java:179)
    at org.apache.hadoop.hive.om.utils.WebHCatAuthenticator$1.run(WebHCatAuthenticator.java:277)
    at org.apache.hadoop.hive.om.utils.WebHCatAuthenticator$1.run(WebHCatAuthenticator.java:253)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.hive.om.utils.WebHCatAuthenticator.doSpnegoSequence(WebHCatAuthenticator.java:253)
    ... 5 more
Caused by: KrbException: Server not found in Kerberos database (7) - UNKNOWN_SERVER
    at sun.security.krb5.KrbTgsRep.<init>(KrbTgsRep.java:73)
    at sun.security.krb5.KrbTgsReq.getReply(KrbTgsReq.java:251)
    at sun.security.krb5.KrbTgsReq.sendAndGetCreds(KrbTgsReq.java:262)
    at sun.security.krb5.internal.CredentialsUtil.serviceCreds(CredentialsUtil.java:308)
    at sun.security.krb5.internal.CredentialsUtil.acquireServiceCreds(CredentialsUtil.java:126)
    at sun.security.krb5.Credentials.acquireServiceCreds(Credentials.java:458)
    at sun.security.jgss.krb5.Krb5Context.initSecContext(Krb5Context.java:693)
    ... 12 more
Caused by: KrbException: Identifier doesn't match expected value (906)
    at sun.security.krb5.internal.KDCRep.init(KDCRep.java:140)
    at sun.security.krb5.internal.TGSRep.init(TGSRep.java:65)
    at sun.security.krb5.internal.TGSRep.<init>(TGSRep.java:60)
    at sun.security.krb5.KrbTgsRep.<init>(KrbTgsRep.java:55)
```

Cause Analysis

1. The server account of the MRS WebHCat role involves the hostname. If you change the hostname after the installation, WebHCat fails to start.
2. The one-to-many or many-to-one association between IP addresses and hostnames is configured in the **/etc/hosts** file. As a result, the IP address and hostname cannot be obtained correctly after the **hostname** and **hostname -i** commands are executed.

Solution

Step 1 Change the hostname of the modified node to the hostname before the cluster is installed.

Step 2 Check whether the **/etc/hosts** of the node where WebHCat is located is correctly configured.

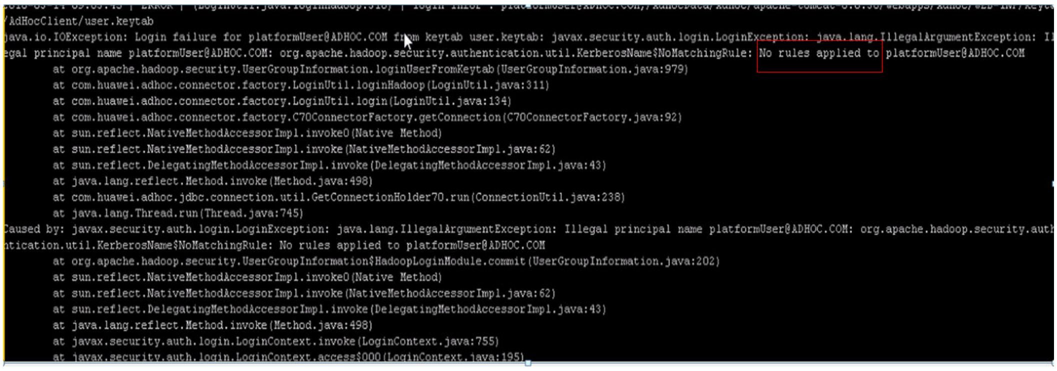
Step 3 Restart WebHCat.

----End

15.10.31 Sample Code Error for Hive Secondary Development After Domain Switching

Symptom

In the sample code for Hive secondary development, an error "No rules applied to ****" is reported:



```
/AdHocClient/user.keytab
java.io.IOException: Login failure for platformUser@ADHOC.COM from keytab user.keytab: javax.security.auth.login.LoginException: java.lang.IllegalArgumentException: Il
egal principal name platformUser@ADHOC.COM: org.apache.hadoop.security.authentication.util.KerberosName$NoMatchingRule: No rules applied to platformUser@ADHOC.COM
        at org.apache.hadoop.security.UserGroupInformation.loginUserFromKeytab(UserGroupInformation.java:979)
        at com.huawei.adhoc.connector.factory.LoginUtil.loginHadoop(LoginUtil.java:311)
        at com.huawei.adhoc.connector.factory.LoginUtil.login(LoginUtil.java:134)
        at com.huawei.adhoc.connector.factory.C70ConnectorFactory.getConnection(C70ConnectorFactory.java:92)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at com.huawei.adhoc.jdbc.connection.util.GetConnectionHolder70.run(ConnectionUtil.java:238)
        at java.lang.Thread.run(Thread.java:745)
Caused by: javax.security.auth.login.LoginException: java.lang.IllegalArgumentException: Illegal principal name platformUser@ADHOC.COM: org.apache.hadoop.security.auth
ntication.util.KerberosName$NoMatchingRule: No rules applied to platformUser@ADHOC.COM
        at org.apache.hadoop.security.UserGroupInformation$HadoopLoginModule.commit(UserGroupInformation.java:202)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at javax.security.auth.login.LoginContext.invoke(LoginContext.java:755)
        at javax.security.auth.login.LoginContext.access$000(LoginContext.java:195)
```

Cause Analysis

1. The sample code for Hive secondary development loads **core-site.xml** file that is loaded through classload by default. Therefore, you need to put the configuration file to the **classpath** directory of the startup program.
2. If the domain name of the cluster is changed, the **core-site.xml** file will change. You need to download the latest **core-site.xml** file and save it to the **classpath** directory where the sample code for Hive secondary development is located.

Solution

Step 1 Download the latest client of the Hive cluster to obtain the latest **core-site.xml** file.

Step 2 Save the **core-site.xml** file to the **classpath** directory where the sample code process for Hive secondary development is located.

----End

15.10.32 MetaStore Exception Occurs When the Number of DBService Connections Exceeds the Upper Limit

Symptom

By default, the maximum number of connections to DBService is 300. If the number of connections is greater than 300 due to heavy traffic, an exception occurs in MetaStore and error "slots are reserved for non-replication superuser connections" is reported.

```
2018-04-26 14:58:55,657 | ERROR | BoneCP-pool-watch-thread | Failed to acquire connection to
jdbc:postgresql://10.*.*.*:20051/hivemeta?socketTimeout=60. Sleeping for 1000 ms. Attempts left: 9 |
com.jolbox.bonecp.BoneCP.obtainInternalConnection(BoneCP.java:292)
org.postgresql.util.PSQLException: FATAL: remaining connection slots are reserved for non-replication
superuser connections
    at org.postgresql.core.v3.ConnectionFactoryImpl.readStartupMessages(ConnectionFactoryImpl.java:643)
    at org.postgresql.core.v3.ConnectionFactoryImpl.openConnectionImpl(ConnectionFactoryImpl.java:184)
    at org.postgresql.core.ConnectionFactory.openConnection(ConnectionFactory.java:64)
    at org.postgresql.jdbc2.AbstractJdbc2Connection.<init>(AbstractJdbc2Connection.java:124)
    at org.postgresql.jdbc3.AbstractJdbc3Connection.<init>(AbstractJdbc3Connection.java:28)
    at org.postgresql.jdbc3g.AbstractJdbc3gConnection.<init>(AbstractJdbc3gConnection.java:20)
    at org.postgresql.jdbc4.AbstractJdbc4Connection.<init>(AbstractJdbc4Connection.java:30)
    at org.postgresql.jdbc4.Jdbc4Connection.<init>(Jdbc4Connection.java:22)
    at org.postgresql.Driver.makeConnection(Driver.java:392)
    at org.postgresql.Driver.connect(Driver.java:266)
    at java.sql.DriverManager.getConnection(DriverManager.java:664)
    at java.sql.DriverManager.getConnection(DriverManager.java:208)
    at com.jolbox.bonecp.BoneCP.obtainRawInternalConnection(BoneCP.java:361)
    at com.jolbox.bonecp.BoneCP.obtainInternalConnection(BoneCP.java:269)
    at com.jolbox.bonecp.ConnectionHandle.<init>(ConnectionHandle.java:242)
    at com.jolbox.bonecp.PoolWatchThread.fillConnections(PoolWatchThread.java:115)
    at com.jolbox.bonecp.PoolWatchThread.run(PoolWatchThread.java:82)
    at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1142)
    at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:617)
    at java.lang.Thread.run(Thread.java:745)
```

Cause Analysis

Heavy service traffic causes more than 300 connections to DBService, and the maximum number of connections to DBService needs to be increased.

Solution

Step 1 Go to the DBService configuration page.

- For versions earlier than MRS 2.0.1: Log in to MRS Manager, choose **Services** > **DBService** > **Service Configuration**, and select **All** from the **Basic** drop-down list.
- For MRS 2.0.1 or later: Click the cluster name on the MRS console, choose **Components** > **DBService** > **Service Configuration**, and select **All** from the **Basic** drop-down list.

  NOTE

  If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)
- For MRS 3.*x* or later: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the target cluster and choose **Services** > **DBService** > **Configurations** > **All Configurations**.

Step 2 Search for **dbservice.database.max.connections** and set it to a proper value not greater than **1000**.

Step 3 Save the configuration and restart the affected services or instances.

Step 4 If the fault persists, check the service code for any connection leaks.

----End

15.10.33 "Failed to execute session hooks: over max connections" Reported by Beeline

Symptom

The default maximum connections to HiveServer are 200. When the number of connections exceeds 200, Beeline reports error "Failed to execute session hooks: over max connections."

```
beeline> [root@172-27-16-38 c70client]# beeline
Connecting to
jdbc:hive2://129.188.82.38:24002,129.188.82.36:24002,129.188.82.35:24002/;serviceDiscoveryMode=zooKeepe
r;zooKeeperNamespace=hiveserver2;sasl.qop=auth-conf;auth=KERBEROS;principal=hive/
hadoop.hadoop.com@HADOOP.COM
Debug is  true storeKey false useTicketCache true useKeyTab false doNotPrompt false ticketCache is null
isInitiator true KeyTab is null refreshKrb5Config is false principal is null tryFirstPass is false useFirstPass is
false storePass is false clearPass is false
Acquire TGT from Cache
Principal is xxx@HADOOP.COM
Commit Succeeded

Error: Failed to execute session hooks: over max connections. (state=,code=0)
Beeline version 1.2.1 by Apache Hive
```

The HiveServer log (**/var/log/Bigdata/hive/hiveserver/hive.log**) shows that error "over max connections" is reported.

```
2018-05-03 04:31:56,728 | WARN  | HiveServer2-Handler-Pool: Thread-137 | Error opening session:  |
org.apache.hive.service.cli.thrift.ThriftCLIService.OpenSession(ThriftCLIService.java:542)
org.apache.hive.service.cli.HiveSQLException: Failed to execute session hooks: over max connections.
    at org.apache.hive.service.cli.session.SessionManager.openSession(SessionManager.java:322)
    at org.apache.hive.service.cli.CLIService.openSessionWithImpersonation(CLIService.java:189)
    at org.apache.hive.service.cli.thrift.ThriftCLIService.getSessionHandle(ThriftCLIService.java:663)
    at org.apache.hive.service.cli.thrift.ThriftCLIService.OpenSession(ThriftCLIService.java:527)
    at org.apache.hive.service.cli.thrift.TCLIService$Processor$OpenSession.getResult(TCLIService.java:1257)
    at org.apache.hive.service.cli.thrift.TCLIService$Processor$OpenSession.getResult(TCLIService.java:1242)
    at org.apache.thrift.ProcessFunction.process(ProcessFunction.java:39)
    at org.apache.thrift.TBaseProcessor.process(TBaseProcessor.java:39)
    at org.apache.hadoop.hive.thrift.HadoopThriftAuthBridge$Server
$TUGIAssumingProcessor.process(HadoopThriftAuthBridge.java:710)
    at org.apache.thrift.server.TThreadPoolServer$WorkerProcess.run(TThreadPoolServer.java:286)
    at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1142)
    at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:617)
    at java.lang.Thread.run(Thread.java:745)
Caused by: org.apache.hive.service.cli.HiveSQLException: over max connections.
    at
org.apache.hadoop.hive.transporthook.SessionControllerTsaslTransportHook.checkTotalSessionNumber(Sessi
onControllerTsaslTransportHook.java:208)
    at
org.apache.hadoop.hive.transporthook.SessionControllerTsaslTransportHook.postOpen(SessionControllerTsasl
TransportHook.java:163)
    at
org.apache.hadoop.hive.transporthook.SessionControllerTsaslTransportHook.run(SessionControllerTsaslTransp
ortHook.java:134)
    at org.apache.hive.service.cli.session.SessionManager.executeSessionHooks(SessionManager.java:432)
    at org.apache.hive.service.cli.session.SessionManager.openSession(SessionManager.java:314)
    ... 12 more
```

Cause Analysis

Heavy service traffic causes the number of connections to one HiveServer node to exceed 200, and the maximum number of connections to HiveServer needs to be increased.

Solution

Step 1 Go to the Hive configuration page.

- For versions earlier than MRS 2.0.1: Log in to MRS Manager, choose **Services** > **Hive** > **Service Configuration**, and select **All** from the **Basic** drop-down list.
- For MRS 2.0.1 or later: Click the cluster name on the MRS console, choose **Components** > **Hive** > **Service Configuration**, and select **All** from the **Basic** drop-down list.

  NOTE

  If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)
- For MRS 3.*x* or later: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the target cluster and choose **Services** > **Hive** > **Configurations** > **All Configurations**.

Step 2 Search for **hive.server.session.control.maxconnections** and set it to a proper value not greater than **1000**.

Step 3 Save the configuration and restart the affected services or instances.

----End

15.10.34 beeline Reports the "OutOfMemoryError" Error

Symptom

When a large amount of data is queried on the Beeline client, the message "OutOFMemoryError: Java heap space" is displayed. The detailed error information is as follows:

```
org.apache.thrift.TException: Error in calling method FetchResults
    at org.apache.hive.jdbc.HiveConnection$SynchronizedHandler.invoke(HiveConnection.java:1514)
    at com.sun.proxy.$Proxy4.FetchResults(Unknown Source)
    at org.apache.hive.jdbc.HiveQueryResultSet.next(HiveQueryResultSet.java:358)
    at org.apache.hive.beeline.BufferedRows.<init>(BufferedRows.java:42)
    at org.apache.hive.beeline.BeeLine.print(BeeLine.java:1856)
    at org.apache.hive.beeline.Commands.execute(Commands.java:873)
    at org.apache.hive.beeline.Commands.sql(Commands.java:714)
    at org.apache.hive.beeline.BeeLine.dispatch(BeeLine.java:1035)
    at org.apache.hive.beeline.BeeLine.execute(BeeLine.java:821)
    at org.apache.hive.beeline.BeeLine.begin(BeeLine.java:778)
    at org.apache.hive.beeline.BeeLine.mainWithInputRedirection(BeeLine.java:486)
    at org.apache.hive.beeline.BeeLine.main(BeeLine.java:469)
Caused by: java.lang.OutOfMemoryError: Java heap space
    at com.sun.crypto.provider.CipherCore.doFinal(CipherCore.java:959)
    at com.sun.crypto.provider.CipherCore.doFinal(CipherCore.java:824)
    at com.sun.crypto.provider.AESCipher.engineDoFinal(AESCipher.java:436)
    at javax.crypto.Cipher.doFinal(Cipher.java:2223)
    at sun.security.krb5.internal.crypto.dk.AesDkCrypto.decryptCTS(AesDkCrypto.java:414)
    at sun.security.krb5.internal.crypto.dk.AesDkCrypto.decryptRaw(AesDkCrypto.java:291)
    at sun.security.krb5.internal.crypto.Aes256.decryptRaw(Aes256.java:86)
    at sun.security.jgss.krb5.CipherHelper.aes256Decrypt(CipherHelper.java:1397)
    at sun.security.jgss.krb5.CipherHelper.decryptData(CipherHelper.java:576)
    at sun.security.jgss.krb5.WrapToken_v2.getData(WrapToken_v2.java:130)
    at sun.security.jgss.krb5.WrapToken_v2.getData(WrapToken_v2.java:105)
    at sun.security.jgss.krb5.Krb5Context.unwrap(Krb5Context.java:1058)
    at sun.security.jgss.GSSContextImpl.unwrap(GSSContextImpl.java:403)
    at com.sun.security.sasl.gsskerb.GssKrb5Base.unwrap(GssKrb5Base.java:77)
    at org.apache.thrift.transport.TSaslTransport$SaslParticipant.unwrap(TSaslTransport.java:559)
    at org.apache.thrift.transport.TSaslTransport.readFrame(TSaslTransport.java:462)
    at org.apache.thrift.transport.TSaslTransport.read(TSaslTransport.java:435)
    at org.apache.thrift.transport.TSaslClientTransport.read(TSaslClientTransport.java:37)
    at org.apache.thrift.transport.TTransport.xxx(TTransport.java:86)
    at org.apache.hadoop.hive.thrift.TFilterTransport.xxx(TFilterTransport.java:62)
    at org.apache.thrift.protocol.TBinaryProtocol.xxx(TBinaryProtocol.java:429)
    at org.apache.thrift.protocol.TBinaryProtocol.readI32(TBinaryProtocol.java:318)
    at org.apache.thrift.protocol.TBinaryProtocol.readMessageBegin(TBinaryProtocol.java:219)
    at org.apache.thrift.TServiceClient.receiveBase(TServiceClient.java:77)
    at org.apache.hive.service.cli.thrift.TCLIService$Client.recv_FetchResults(TCLIService.java:505)
    at org.apache.hive.service.cli.thrift.TCLIService$Client.FetchResults(TCLIService.java:492)
    at sun.reflect.GeneratedMethodAccessor2.invoke(Unknown Source)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at org.apache.hive.jdbc.HiveConnection$SynchronizedHandler.invoke(HiveConnection.java:1506)
    at com.sun.proxy.$Proxy4.FetchResults(Unknown Source)
    at org.apache.hive.jdbc.HiveQueryResultSet.next(HiveQueryResultSet.java:358)
Error: Error retrieving next row (state=,code=0)
```

Cause Analysis

- The data volume is excessively large.
- Users use the **select * from table_name;** statement for query in the whole table. There is a large amount of data in the table.
- The default startup memory of Beeline is 128 MB. The returned result set is too large during query, overloading Beeline.

Solution

Step 1 Before running **select count(*) from table_name;**, check the amount of data to be queried and determine whether to display data of this magnitude in Beeline.

Step 2 If a certain amount of data needs to be displayed, adjust the JVM parameter of the Hive client. Add **export HIVE_OPTS=-Xmx1024M** (change the value based on service requirements) to **component_env** in the **/Hive** directory of the Hive client. Run the **source** command to obtain the **/bigdata_env** directory on the client.

----End

15.10.35 Task Execution Fails Because the Input File Number Exceeds the Threshold

Symptom

When Hive performs a query operation, error message "Job Submission failed with exception 'java.lang.RuntimeException(input file number exceeded the limits in the conf;input file num is: 2380435,max heap memory is: 16892035072,the limit conf is: 500000/4)'" is displayed. The value in the error message varies depending on the actual situation. The error details are as follows:

```
ERROR : Job Submission failed with exception 'java.lang.RuntimeException(input file numbers exceeded the
limits in the conf;
 input file num is: 2380435 ,
 max heap memory is: 16892035072 ,
 the limit conf is: 500000/4)'
java.lang.RuntimeException: input file numbers exceeded the limits in the conf;
 input file num is: 2380435 ,
 max heap memory is: 16892035072 ,
 the limit conf is: 500000/4
    at org.apache.hadoop.hive.ql.exec.mr.ExecDriver.checkFileNum(ExecDriver.java:545)
    at org.apache.hadoop.hive.ql.exec.mr.ExecDriver.execute(ExecDriver.java:430)
    at org.apache.hadoop.hive.ql.exec.mr.MapRedTask.execute(MapRedTask.java:137)
    at org.apache.hadoop.hive.ql.exec.Task.executeTask(Task.java:158)
    at org.apache.hadoop.hive.ql.exec.TaskRunner.runSequential(TaskRunner.java:101)
    at org.apache.hadoop.hive.ql.Driver.launchTask(Driver.java:1965)
    at org.apache.hadoop.hive.ql.Driver.execute(Driver.java:1723)
    at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1475)
    at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1283)
    at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1278)
    at org.apache.hive.service.cli.operation.SQLOperation.runQuery(SQLOperation.java:167)
    at org.apache.hive.service.cli.operation.SQLOperation.access$200(SQLOperation.java:75)
    at org.apache.hive.service.cli.operation.SQLOperation$1$1.run(SQLOperation.java:245)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1710)
    at org.apache.hive.service.cli.operation.SQLOperation$1.run(SQLOperation.java:258)
    at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:511)
    at java.util.concurrent.FutureTask.run(FutureTask.java:266)
    at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1142)
    at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:617)
    at java.lang.Thread.run(Thread.java:745)

Error: Error while processing statement: FAILED: Execution Error, return code 1 from
org.apache.hadoop.hive.ql.exec.mr.MapRedTask (state=08S01,code=1)
```

Cause Analysis

MRS uses the ratio of maximum files to the maximum HiveServer heap memory to determine the number of input files allowed in a MapReduce job submission. Default value **500000/4** indicates that each 4 GB of heap memory allows a maximum of 500,000 input files. An error occurs if the number of input files exceeds this limit.

Solution

Step 1 Go to the Hive configuration page.

- For versions earlier than MRS 2.0.1: Log in to MRS Manager, choose **Services** > **Hive** > **Service Configuration**, and select **All** from the **Basic** drop-down list.
- For MRS 2.0.1 or later: Click the cluster name on the MRS console, choose **Components** > **Hive** > **Service Configuration**, and select **All** from the **Basic** drop-down list.

  NOTE

  If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

- For MRS 3.*x* or later: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the target cluster and choose **Services** > **Hive** > **Configurations** > **All Configurations**.

----End

Step 1 Search for **hive.mapreduce.input.files2memory** and set it to a proper value based on the actual memory and task.

Step 2 Save the configuration and restart the affected services or instances.

Step 3 If the fault persists, adjust the GC parameter of the HiveServer based on service requirements.

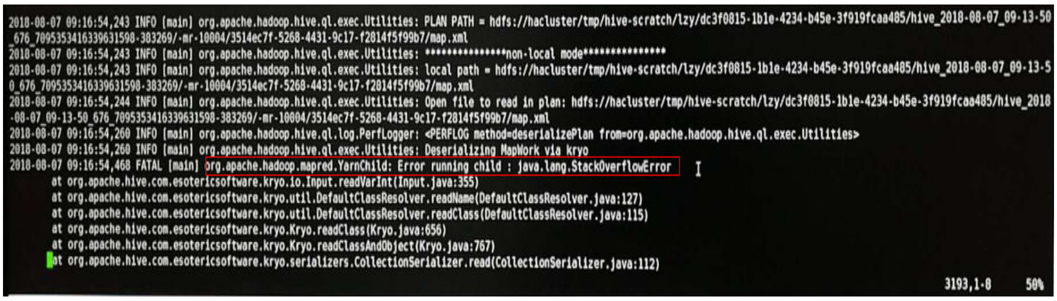
----End

15.10.36 Task Execution Fails Because of Stack Memory Overflow

Symptom

When Hive performs a query operation, error "Error running child: java.lang.StackOverflowError" is reported. The error details are as follows:

```
FATAL [main] org.apache.hadoop.mapred.YarnChild: Error running child : java.lang.StackOverflowError
at org.apache.hive.come.esotericsoftware.kryo.io.Input.readVarInt(Input.java:355)
at org.apache.hive.come.esotericsoftware.kryo.util.DefautClassResolver.readName(DefautClassResolver.java:
127)
at org.apache.hive.come.esotericsoftware.kryo.util.DefautClassResolver.readClass(DefautClassResolver.java:
115)
at org.apache.hive.come.esotericsoftware.kryo.Kryo.readClass(Kryo.java.656)
at org.apache.hive.come.esotericsoftware.kryo.kryo.readClassAnd0bject(Kryo.java:767)
at org.apache.hive.come.esotericsoftware.kryo.serializers.collectionSerializer.read(CollectionSerializer.java:
112)
```



```
2018-08-07 09:16:54,243 INFO [main] org.apache.hadoop.hive.ql.exec.Utilities: PLAN PATH = hdfs://hacluster/tmp/hive-scratch/lzy/dc3f0815-1b1e-4234-b45e-3f919fcaa485/hive_2018-08-07_09-13-50
_676_7095353416339631598-383269/-mr-10004/3514ec7f-5268-4431-9c17-f2814f5f99b7/map.xml
2018-08-07 09:16:54,243 INFO [main] org.apache.hadoop.hive.ql.exec.Utilities: ***************non-local mode***************
2018-08-07 09:16:54,243 INFO [main] org.apache.hadoop.hive.ql.exec.Utilities: local path = hdfs://hacluster/tmp/hive-scratch/lzy/dc3f0815-1b1e-4234-b45e-3f919fcaa485/hive_2018-08-07_09-13-5
0_676_7095353416339631598-383269/-mr-10004/3514ec7f-5268-4431-9c17-f2814f5f99b7/map.xml
2018-08-07 09:16:54,244 INFO [main] org.apache.hadoop.hive.ql.exec.Utilities: Open file to read in plan: hdfs://hacluster/tmp/hive-scratch/lzy/dc3f0815-1b1e-4234-b45e-3f919fcaa485/hive_2018
-08-07_09-13-50_676_7095353416339631598-383269/-mr-10004/3514ec7f-5268-4431-9c17-f2814f5f99b7/map.xml
2018-08-07 09:16:54,260 INFO [main] org.apache.hadoop.hive.ql.log.PerfLogger: <PERFLOG method=deserializePlan from=org.apache.hadoop.hive.ql.exec.Utilities>
2018-08-07 09:16:54,260 INFO [main] org.apache.hadoop.hive.ql.exec.Utilities: Deserializing MapWork via kryo
2018-08-07 09:16:54,468 FATAL [main] org.apache.hadoop.mapred.YarnChild: Error running child : java.lang.StackOverflowError
        at org.apache.hive.com.esotericsoftware.kryo.io.Input.readVarInt(Input.java:355)
        at org.apache.hive.com.esotericsoftware.kryo.util.DefaultClassResolver.readName(DefaultClassResolver.java:127)
        at org.apache.hive.com.esotericsoftware.kryo.util.DefaultClassResolver.readClass(DefaultClassResolver.java:115)
        at org.apache.hive.com.esotericsoftware.kryo.Kryo.readClass(Kryo.java:656)
        at org.apache.hive.com.esotericsoftware.kryo.Kryo.readClassAndObject(Kryo.java:767)
        at org.apache.hive.com.esotericsoftware.kryo.serializers.CollectionSerializer.read(CollectionSerializer.java:112)
                                                                                                    3193,1-8      50%
```

Cause Analysis

Error "java.lang.StackOverflowError" indicates the memory overflow of the thread stack. It may occur if there are multiple levels of calls (for example, infinite recursive calls) or the thread stack is too small.

Solution

Adjust the stack memory in the JVM parameters of the Map and Reduce stages during execution of a MapReduce job, that is, **mapreduce.map.java.opts** (adjusting the stack memory of Map) and **mapreduce.reduce.java.opts** (adjusting the stack memory of Reduce). The following uses the **mapreduce.map.java.opts** parameter as an example.

- To increase the Map memory temporarily (only valid for Beeline):

  Run the **set mapreduce.map.java.opts=-Xss8G;** command on the Beeline client. (Change the value as required.)

- To permanently increase the Map memory specified by the **mapreduce.map.memory.mb** and **mapreduce.map.java.opts** parameters:

  a. Go to the Hive configuration page.

     - For versions earlier than MRS 2.0.1: Log in to MRS Manager, choose **Services** > **Hive** > **Service Configuration**, and select **All** from the **Basic** drop-down list.

     - For MRS 2.0.1 or later: Click the cluster name on the MRS console, choose **Components** > **Hive** > **Service Configuration**, and select **All** from the **Basic** drop-down list.

       NOTE

       If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)

     - For MRS 3.*x* or later: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the target cluster and choose **Services** > **Hive** > **Configurations** > **All Configurations**.

  b. Add custom parameter **mapreduce.map.java.opts** and set it to a proper value.

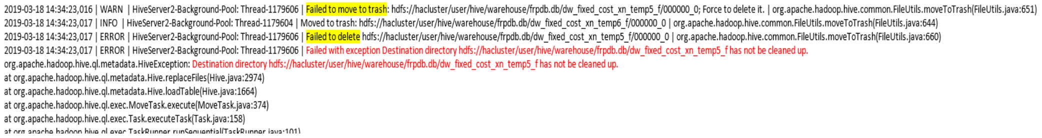
  c. Save the configuration and restart the affected services or instances.

     Note that the modification takes effect after a service restart. During the restart, the Hive service is unavailable.

15.10.37 Task Failed Due to Concurrent Writes to One Table or Partition

Symptom

When Hive executes an INSERT statement, an error is reported indicating that a file or directory already exists or is cleared in HDFS. The error details are as follows:



```
2019-03-18 14:34:23,016 | WARN  | HiveServer2-Background-Pool: Thread-1179606 | Failed to move to trash: hdfs://hacluster/user/hive/warehouse/frpdb.db/dw_fixed_cost_xn_temp5_f/000000_0; Force to delete it. | org.apache.hadoop.hive.common.FileUtils.moveToTrash(FileUtils.java:651)
2019-03-18 14:34:23,017 | INFO  | HiveServer2-Background-Pool: Thread-1179604 | Moved to trash: hdfs://hacluster/user/hive/warehouse/frpdb.db/dw_fixed_cost_xn_temp6_f/000000_0 | org.apache.hadoop.hive.common.FileUtils.moveToTrash(FileUtils.java:644)
2019-03-18 14:34:23,017 | ERROR | HiveServer2-Background-Pool: Thread-1179606 | Failed to delete hdfs://hacluster/user/hive/warehouse/frpdb.db/dw_fixed_cost_xn_temp5_f/000000_0 | org.apache.hadoop.hive.common.FileUtils.moveToTrash(FileUtils.java:660)
2019-03-18 14:34:23,017 | ERROR | HiveServer2-Background-Pool: Thread-1179606 | Failed with exception Destination directory hdfs://hacluster/user/hive/warehouse/frpdb.db/dw_fixed_cost_xn_temp5_f has not be cleaned up.
org.apache.hadoop.hive.ql.metadata.HiveException: Destination directory hdfs://hacluster/user/hive/warehouse/frpdb.db/dw_fixed_cost_xn_temp5_f has not be cleaned up.
at org.apache.hadoop.hive.ql.metadata.Hive.replaceFiles(Hive.java:2974)
at org.apache.hadoop.hive.ql.metadata.Hive.loadTable(Hive.java:1664)
at org.apache.hadoop.hive.ql.exec.MoveTask.execute(MoveTask.java:374)
at org.apache.hadoop.hive.ql.exec.Task.executeTask(Task.java:158)
at org.apache.hadoop.hive.ql.exec.TaskRunner.runSequential(TaskRunner.java:101)
```

Cause Analysis

1. Check the start time and end time of the task based on the HiveServer audit logs.
2. Check whether data is inserted into the same table or partition in the time segment.

3. Hive does not support concurrent data insertion for a table or partition. As a result, multiple tasks perform operations on the same temporary data directory, and one task moves the data of another task, causing task failure.

Solution

The service logic is modified so that data is inserted to the same table or partition in single thread mode.

15.10.38 Hive Task Failed Due to a Lack of HDFS Directory Permission

Symptom

An error message is displayed, indicating that the user does not have the permission to access the HDFS directory.

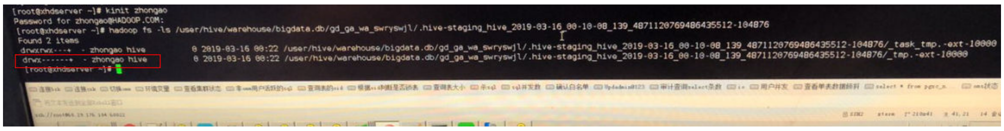
```
2019-04-09 17:49:19,845 | ERROR | HiveServer2-Background-Pool: Thread-3160445 | Job Submission failed
with exception 'org.apache.hadoop.security.AccessControlException(Permission denied: user=hive_quanxian,
access=READ_EXECUTE, inode="/user/hive/warehouse/bigdata.db/
gd_ga_wa_swryswjl":zhongao:hive:drwx------
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkAccessAcl(FSPermissionChecker.java:
426)
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.check(FSPermissionChecker.java:329)
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkSubAccess(FSPermissionChecker.java:
300)
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkPermission(FSPermissionChecker.java:
241)
at
com.xxx.hadoop.adapter.hdfs.plugin.HWAccessControlEnforce.checkPermission(HWAccessControlEnforce.java:
69)
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkPermission(FSPermissionChecker.java:
190)
at org.apache.hadoop.hdfs.server.namenode.FSDirectory.checkPermission(FSDirectory.java:1910)
at org.apache.hadoop.hdfs.server.namenode.FSDirectory.checkPermission(FSDirectory.java:1894)
at
org.apache.hadoop.hdfs.server.namenode.FSDirStatAndListingOp.getContentSummary(FSDirStatAndListingO
p.java:135)
at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getContentSummary(FSNamesystem.java:3983)
at
org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.getContentSummary(NameNodeRpcServer.ja
va:1342)
at
org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolServerSideTranslatorPB.getContentSummary(Cli
entNamenodeProtocolServerSideTranslatorPB.java:925)
at org.apache.hadoop.hdfs.protocol.proto.ClientNamenodeProtocolProtos$ClientNamenodeProtocol
$2.callBlockingMethod(ClientNamenodeProtocolProtos.java)
at org.apache.hadoop.ipc.ProtobufRpcEngine$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine.java:616)
at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:973)
at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2260)
at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2256)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:422)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1781)
at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2254)
)'
```

Cause Analysis

1. According to the stack information, the permission on the subdirectory fails to be checked.

```
org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkSubAccess(FSPermissionChecker.ja
va:300)
```

2. Check the permission of all files and directories in HDFS. The permission of a directory is 700 (only the file owner can access the directory). It is confirmed that an abnormal directory exists.



```
[root@xhdserver ~]# kinit zhongao
Password for zhongao@HADOOP.COM:
[root@xhdserver ~]# hadoop fs -ls /user/hive/warehouse/bigdata.db/gd_ga_wa_swryswjl/.hive-staging_hive_2019-03-16_00-10-08_139_4871120769486435512-104876
Found 2 items
drwxrwx---+   - zhongao hive          0 2019-03-16 00:22 /user/hive/warehouse/bigdata.db/gd_ga_wa_swryswjl/.hive-staging_hive_2019-03-16_00-10-08_139_4871120769486435512-104876/_task_tmp.-ext-10000
drwx------+   - zhongao hive          0 2019-03-16 00:22 /user/hive/warehouse/bigdata.db/gd_ga_wa_swryswjl/.hive-staging_hive_2019-03-16_00-10-08_139_4871120769486435512-104876/_tmp.-ext-10000
[root@xhdserver ~]#
```

Solution

1. Check whether the file is imported manually. If not, delete the file or directory.
2. If the file or directory cannot be deleted, change the file or directory permission to 770.

15.10.39 Failed to Load Data to Hive Tables

Symptom

After creating a table, a user runs the **LOAD** command to import data to the table. However, the following problem occurs during the import:

```
.......
> LOAD DATA INPATH '/user/tester1/hive-data/data.txt' INTO TABLE employees_info;
Error: Error while compiling statement: FAILED: SemanticException Unable to load data to destination table.
Error: The file that you are trying to load does not match the file format of the destination table.
(state=42000,code=40000)
..........
```

Cause Analysis

1. The storage format is not specified during table creation, and the default format RCFile is used.
2. However, the data to be imported is in TEXTFILE format.

Solution

This problem is caused by an application defect. You can use a proper method based on site requirements only by ensuring that the storage format specified by the table is the same as the format of the data to be imported.

- Method 1:

  Specify the storage format when creating a table as a user who has the Hive table operation permission. For example:

  CREATE TABLE IF NOT EXISTS employees_info(name STRING,age INT) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE;

  Specify the format of the data to be imported as TEXTFILE.
- Method 2:

  Import RCFile data, but not TEXTFILE data.

15.10.40 HiveServer and HiveHCat Process Faults

Issue

The HiveServer and WebHCat processes in the customer cluster are faulty.

Symptom

The HiveServer and WebHCat processes on the Master2 node in the MRS cluster are faulty. After the restart, the processes are still faulty.

Cause Analysis

On Manager, start the faulty HiveServer process. Log in to the background and search for the error information at the corresponding time point in the **hiveserver.out** log file. The error information is as follows: **error parsing conf mapred-site.xml** and **Premature end of file**. Restart WebHCat. The same error is reported because the **mapred-site.xml** file fails to be parsed.

Procedure

1. Log in to the Master2 node as user **root**.
2. Run the **find / -name 'mapred-site.xml'** command to obtain the location of the **mapred-site.xml** file.
   - The path of HiveServer is **/opt/Bigdata/***Cluster version***/ 1_13_HiveServer/etc/mapred-site.xml**.
   - The path of WebHCat is **/opt/Bigdata/***Cluster version***/ 1_13_WebHCat/etc/mapred-site.xml**.
3. Check whether the **mapred-site.xml** file is normal. In this case, the configuration file is empty. As a result, the parsing fails.
4. Restore the **mapred-site.xml** file. Run the **scp** command to copy the configuration file in the corresponding directory on the Master1 node to the corresponding directory on the Master2 node to replace the original file.
5. Run the **chown omm:wheel mapred-site.xml** command to change the owner group and user.
6. On Manager, restart the faulty HiveServer and WebHCat processes.

15.10.41 An Error Occurs When the INSERT INTO Statement Is Executed on Hive But the Error Message Is Unclear

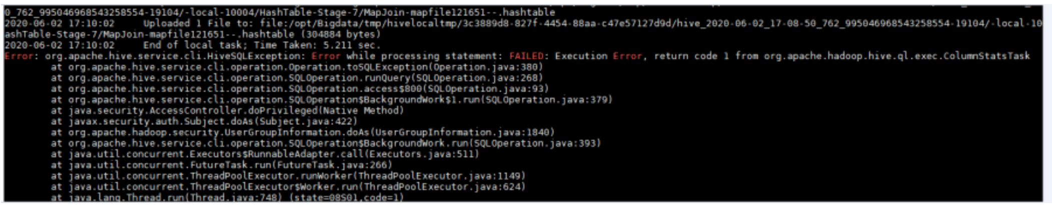
Issue

An error is reported when a user uses MRS Hive to execute a SQL statement.

Symptom

When a user uses MRS Hive to execute a SQL statement, the following error message is displayed.

Figure 15-38 Error reported when MRS Hive executes a SQL statement

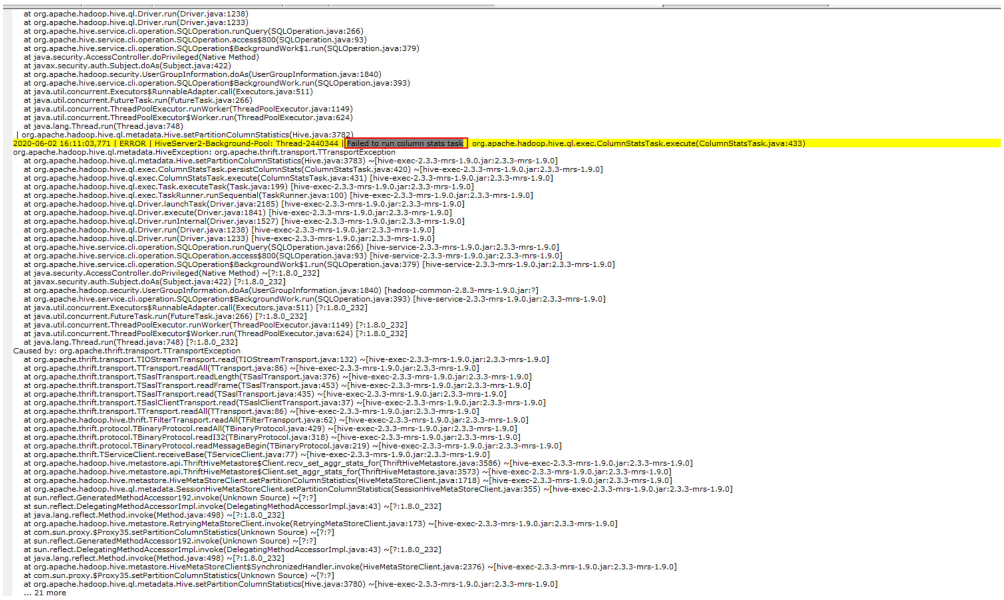


```
0_762_995046968543258554-19104/-local-10004/HashTable-Stage-7/MapJoin-mapfile121651--.hashtable
2020-06-02 17:10:02     Uploaded 1 File to: file:/opt/Bigdata/tmp/hivelocaltmp/3c3889d8-827f-4454-88aa-c47e57127d9d/hive_2020-06-02_17-08-50_762_995046968543258554-19104/-local-10
ashTable-Stage-7/MapJoin-mapfile121651--.hashtable (304884 bytes)
2020-06-02 17:10:02     End of local task; Time Taken: 5.211 sec.
Error: org.apache.hive.service.cli.HiveSQLException: Error while processing statement: FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.ColumnStatsTask
        at org.apache.hive.service.cli.operation.Operation.toSQLException(Operation.java:380)
        at org.apache.hive.service.cli.operation.SQLOperation.runQuery(SQLOperation.java:268)
        at org.apache.hive.service.cli.operation.SQLOperation.access$800(SQLOperation.java:93)
        at org.apache.hive.service.cli.operation.SQLOperation$BackgroundWork$1.run(SQLOperation.java:379)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1840)
        at org.apache.hive.service.cli.operation.SQLOperation$BackgroundWork.run(SQLOperation.java:393)
        at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:511)
        at java.util.concurrent.FutureTask.run(FutureTask.java:266)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
        at java.lang.Thread.run(Thread.java:748) (state=08S01,code=1)
```

Cause Analysis

1. The HiveServer log shows the following message at the time when the error is reported.

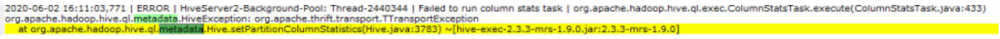
Figure 15-39 HiveServer log



```
   at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1238)
   at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1233)
   at org.apache.hive.service.cli.operation.SQLOperation.runQuery(SQLOperation.java:266)
   at org.apache.hive.service.cli.operation.SQLOperation.access$800(SQLOperation.java:93)
   at org.apache.hive.service.cli.operation.SQLOperation$BackgroundWork$1.run(SQLOperation.java:379)
   at java.security.AccessController.doPrivileged(Native Method)
   at javax.security.auth.Subject.doAs(Subject.java:422)
   at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1840)
   at org.apache.hive.service.cli.operation.SQLOperation$BackgroundWork.run(SQLOperation.java:393)
   at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:511)
   at java.util.concurrent.FutureTask.run(FutureTask.java:266)
   at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)
   at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
   at java.lang.Thread.run(Thread.java:748)
 | org.apache.hadoop.hive.ql.metadata.Hive.setPartitionColumnStatistics(Hive.java:3782)
2020-06-02 16:11:03,771 | ERROR | HiveServer2-Background-Pool: Thread-2440344 | Failed to run column stats task  org.apache.hadoop.hive.ql.exec.ColumnStatsTask.execute(ColumnStatsTask.java:433)
org.apache.hadoop.hive.ql.metadata.HiveException: org.apache.thrift.transport.TTransportException
   at org.apache.hadoop.hive.ql.metadata.Hive.setPartitionColumnStatistics(Hive.java:3783) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.ql.exec.ColumnStatsTask.persistColumnStats(ColumnStatsTask.java:420) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.ql.exec.ColumnStatsTask.execute(ColumnStatsTask.java:431) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.ql.exec.Task.executeTask(Task.java:199) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.ql.exec.TaskRunner.runSequential(TaskRunner.java:100) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.ql.Driver.launchTask(Driver.java:2185) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.ql.Driver.execute(Driver.java:1841) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1527) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1238) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1233) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hive.service.cli.operation.SQLOperation.runQuery(SQLOperation.java:266) [hive-service-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hive.service.cli.operation.SQLOperation.access$800(SQLOperation.java:93) [hive-service-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hive.service.cli.operation.SQLOperation$BackgroundWork$1.run(SQLOperation.java:379) [hive-service-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at java.security.AccessController.doPrivileged(Native Method) ~[?:1.8.0_232]
   at javax.security.auth.Subject.doAs(Subject.java:422) [?:1.8.0_232]
   at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1840) [hadoop-common-2.8.3-mrs-1.9.0.jar:?]
   at org.apache.hive.service.cli.operation.SQLOperation$BackgroundWork.run(SQLOperation.java:393) [hive-service-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:511) [?:1.8.0_232]
   at java.util.concurrent.FutureTask.run(FutureTask.java:266) [?:1.8.0_232]
   at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149) [?:1.8.0_232]
   at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624) [?:1.8.0_232]
   at java.lang.Thread.run(Thread.java:748) [?:1.8.0_232]
Caused by: org.apache.thrift.transport.TTransportException
   at org.apache.thrift.transport.TIOStreamTransport.read(TIOStreamTransport.java:132) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.thrift.transport.TTransport.readAll(TTransport.java:86) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.thrift.transport.TSaslTransport.readLength(TSaslTransport.java:376) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.thrift.transport.TSaslTransport.readFrame(TSaslTransport.java:453) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.thrift.transport.TSaslTransport.read(TSaslTransport.java:435) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.thrift.transport.TSaslClientTransport.read(TSaslClientTransport.java:37) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.thrift.transport.TTransport.readAll(TTransport.java:86) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.thrift.TFilterTransport.readAll(TFilterTransport.java:62) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.thrift.protocol.TBinaryProtocol.readAll(TBinaryProtocol.java:429) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.thrift.protocol.TBinaryProtocol.readI32(TBinaryProtocol.java:318) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.thrift.protocol.TBinaryProtocol.readMessageBegin(TBinaryProtocol.java:219) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.thrift.TServiceClient.receiveBase(TServiceClient.java:77) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.metastore.api.ThriftHiveMetastore$Client.recv_set_aggr_stats_for(ThriftHiveMetastore.java:3586) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.metastore.api.ThriftHiveMetastore$Client.set_aggr_stats_for(ThriftHiveMetastore.java:3573) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.metastore.HiveMetaStoreClient.setPartitionColumnStatistics(HiveMetaStoreClient.java:1718) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.ql.metadata.SessionHiveMetaStoreClient.setPartitionColumnStatistics(SessionHiveMetaStoreClient.java:355) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at sun.reflect.GeneratedMethodAccessor192.invoke(Unknown Source) ~[?:?]
   at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43) ~[?:1.8.0_232]
   at java.lang.reflect.Method.invoke(Method.java:498) ~[?:1.8.0_232]
   at org.apache.hadoop.hive.metastore.RetryingMetaStoreClient.invoke(RetryingMetaStoreClient.java:173) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at com.sun.proxy.$Proxy35.setPartitionColumnStatistics(Unknown Source) ~[?:?]
   at sun.reflect.GeneratedMethodAccessor192.invoke(Unknown Source) ~[?:?]
   at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43) ~[?:1.8.0_232]
   at java.lang.reflect.Method.invoke(Method.java:498) ~[?:1.8.0_232]
   at org.apache.hadoop.hive.metastore.HiveMetaStoreClient$SynchronizedHandler.invoke(HiveMetaStoreClient.java:2376) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at com.sun.proxy.$Proxy35.setPartitionColumnStatistics(Unknown Source) ~[?:?]
   at org.apache.hadoop.hive.ql.metadata.Hive.setPartitionColumnStatistics(Hive.java:3780) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   ... 21 more
```

2. No important information is found in that log, but the **metadata** field is found in the stack. Therefore, the error may be related to MetaStore.

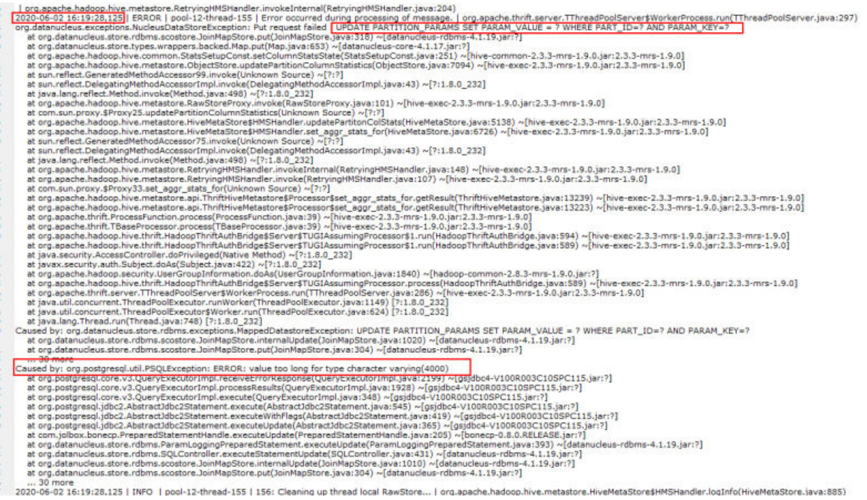
Figure 15-40 Metadata in the stack



```
2020-06-02 16:11:03,771 | ERROR | HiveServer2-Background-Pool: Thread-2440344 | Failed to run column stats task | org.apache.hadoop.hive.ql.exec.ColumnStatsTask.execute(ColumnStatsTask.java:433)
org.apache.hadoop.hive.ql.metadata.HiveException: org.apache.thrift.transport.TTransportException
   at org.apache.hadoop.hive.ql.metadata.Hive.setPartitionColumnStatistics(Hive.java:3783) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
```

3. The MetaStore log shows the following error information.

Figure 15-41 MetaStore log



```
 | org.apache.hadoop.hive.metastore.RetryingHMSHandler.invokeInternal(RetryingHMSHandler.java:204)
2020-06-02 16:19:28,125 | ERROR | pool-12-thread-155 | Error occurred during processing of message. | org.apache.thrift.server.TThreadPoolServer$WorkerProcess.run(TThreadPoolServer.java:297)
org.datanucleus.exceptions.NucleusDataStoreException: Put request failed  UPDATE PARTITION_PARAMS SET PARAM_VALUE = ? WHERE PART_ID=? AND PARAM_KEY=?
   at org.datanucleus.store.rdbms.scostore.JoinMapStore.put(JoinMapStore.java:318) ~[datanucleus-rdbms-4.1.19.jar:?]
   at org.datanucleus.store.types.wrappers.backed.Map.put(Map.java:653) ~[datanucleus-core-4.1.17.jar:?]
   at org.apache.hadoop.hive.common.StatsSetupConst.setColumnStatsState(StatsSetupConst.java:251) ~[hive-common-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.metastore.ObjectStore.updatePartitionColumnStatistics(ObjectStore.java:7094) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at sun.reflect.GeneratedMethodAccessor99.invoke(Unknown Source) ~[?:?]
   at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43) ~[?:1.8.0_232]
   at java.lang.reflect.Method.invoke(Method.java:498) ~[?:1.8.0_232]
   at org.apache.hadoop.hive.metastore.RawStoreProxy.invoke(RawStoreProxy.java:101) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at com.sun.proxy.$Proxy25.updatePartitionColumnStatistics(Unknown Source) ~[?:?]
   at org.apache.hadoop.hive.metastore.HiveMetaStore$HMSHandler.updatePartitonColStats(HiveMetaStore.java:5138) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.metastore.HiveMetaStore$HMSHandler.set_aggr_stats_for(HiveMetaStore.java:6726) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at sun.reflect.GeneratedMethodAccessor75.invoke(Unknown Source) ~[?:?]
   at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43) ~[?:1.8.0_232]
   at java.lang.reflect.Method.invoke(Method.java:498) ~[?:1.8.0_232]
   at org.apache.hadoop.hive.metastore.RetryingHMSHandler.invokeInternal(RetryingHMSHandler.java:148) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.metastore.RetryingHMSHandler.invoke(RetryingHMSHandler.java:107) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at com.sun.proxy.$Proxy33.set_aggr_stats_for(Unknown Source) ~[?:?]
   at org.apache.hadoop.hive.metastore.api.ThriftHiveMetastore$Processor$set_aggr_stats_for.getResult(ThriftHiveMetastore.java:13239) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.metastore.api.ThriftHiveMetastore$Processor$set_aggr_stats_for.getResult(ThriftHiveMetastore.java:13223) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.thrift.ProcessFunction.process(ProcessFunction.java:39) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.thrift.TBaseProcessor.process(TBaseProcessor.java:39) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.thrift.HadoopThriftAuthBridge$Server$TUGIAssumingProcessor$1.run(HadoopThriftAuthBridge.java:594) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.hadoop.hive.thrift.HadoopThriftAuthBridge$Server$TUGIAssumingProcessor$1.run(HadoopThriftAuthBridge.java:589) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at java.security.AccessController.doPrivileged(Native Method) ~[?:1.8.0_232]
   at javax.security.auth.Subject.doAs(Subject.java:422) ~[?:1.8.0_232]
   at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1840) ~[hadoop-common-2.8.3-mrs-1.9.0.jar:?]
   at org.apache.hadoop.hive.thrift.HadoopThriftAuthBridge$Server$TUGIAssumingProcessor.process(HadoopThriftAuthBridge.java:589) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at org.apache.thrift.server.TThreadPoolServer$WorkerProcess.run(TThreadPoolServer.java:286) ~[hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
   at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149) [?:1.8.0_232]
   at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624) [?:1.8.0_232]
   at java.lang.Thread.run(Thread.java:748) [?:1.8.0_232]
Caused by: org.datanucleus.store.rdbms.exceptions.MappedDatastoreException: UPDATE PARTITION_PARAMS SET PARAM_VALUE = ? WHERE PART_ID=? AND PARAM_KEY=?
   at org.datanucleus.store.rdbms.scostore.JoinMapStore.internalUpdate(JoinMapStore.java:1020) ~[datanucleus-rdbms-4.1.19.jar:?]
   at org.datanucleus.store.rdbms.scostore.JoinMapStore.put(JoinMapStore.java:304) ~[datanucleus-rdbms-4.1.19.jar:?]
   ... 30 more
Caused by: org.postgresql.util.PSQLException: ERROR: value too long for type character varying(4000)
   at org.postgresql.core.v3.QueryExecutorImpl.receiveErrorResponse(QueryExecutorImpl.java:2199) ~[gsjdbc4-V100R003C10SPC115.jar:?]
   at org.postgresql.core.v3.QueryExecutorImpl.processResults(QueryExecutorImpl.java:1928) ~[gsjdbc4-V100R003C10SPC115.jar:?]
   at org.postgresql.core.v3.QueryExecutorImpl.execute(QueryExecutorImpl.java:348) ~[gsjdbc4-V100R003C10SPC115.jar:?]
   at org.postgresql.jdbc2.AbstractJdbc2Statement.execute(AbstractJdbc2Statement.java:545) ~[gsjdbc4-V100R003C10SPC115.jar:?]
   at org.postgresql.jdbc2.AbstractJdbc2Statement.executeWithFlags(AbstractJdbc2Statement.java:419) ~[gsjdbc4-V100R003C10SPC115.jar:?]
   at org.postgresql.jdbc2.AbstractJdbc2Statement.executeUpdate(AbstractJdbc2Statement.java:365) ~[gsjdbc4-V100R003C10SPC115.jar:?]
   at com.jolbox.bonecp.PreparedStatementHandle.executeUpdate(PreparedStatementHandle.java:205) ~[bonecp-0.8.0.RELEASE.jar:?]
   at org.datanucleus.store.rdbms.ParamLoggingPreparedStatement.executeUpdate(ParamLoggingPreparedStatement.java:393) ~[datanucleus-rdbms-4.1.19.jar:?]
   at org.datanucleus.store.rdbms.SQLController.executeStatementUpdate(SQLController.java:431) ~[datanucleus-rdbms-4.1.19.jar:?]
   at org.datanucleus.store.rdbms.scostore.JoinMapStore.internalUpdate(JoinMapStore.java:1010) ~[datanucleus-rdbms-4.1.19.jar:?]
   at org.datanucleus.store.rdbms.scostore.JoinMapStore.put(JoinMapStore.java:304) ~[datanucleus-rdbms-4.1.19.jar:?]
   ... 30 more
2020-06-02 16:19:28,125 | INFO  | pool-12-thread-155 | 156: Cleaning up thread local RawStore... | org.apache.hadoop.hive.metastore.HiveMetaStore$HMSHandler.logInfo(HiveMetaStore.java:885)
```

The error context indicates that an error occurs during SQL statement execution, and the following information is displayed in the error message:

```
Caused by: org.postgresql.util.PSQLException: ERROR: value too long for type character varying(4000)
```

The SQL statement fails because the length of all columns exceeds 4000 bytes. The restriction needs to be modified.

Procedure

Step 1 Log in to any master node in the cluster as user **root** and run the **su - omm** command to switch to user **omm**.

Step 2 Run the following command to log in to GaussDB:

gsql -p 20051 -d hivemeta -U *username* -W *password*

Step 3 Run the following command to modify the restriction:

alter table PARTITION_PARAMS alter column PARAM_VALUE type varchar(6000);

----End

15.10.42 Timeout Reported When Adding the Hive Table Field

Issue

An error message is reported when adding the Hive table fields.

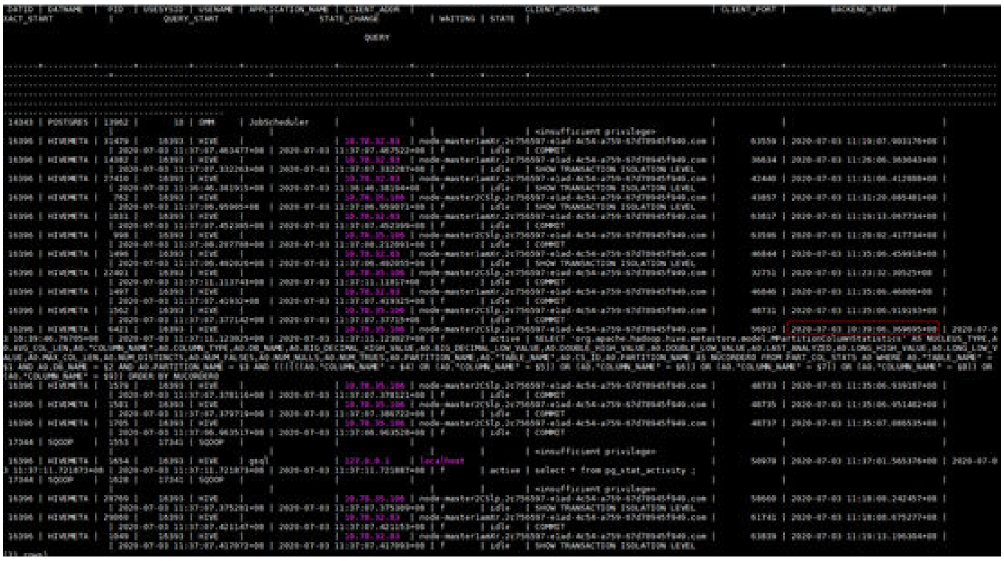
Symptom

Hive executes **ALTER TABLE table_name ADD COLUMNS(column_name string) CASCADE** on tables that contain more than 10,000 partitions. The error information is as follows:

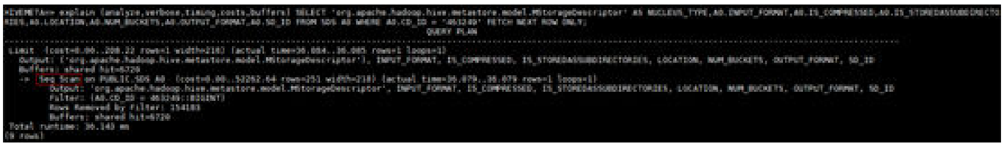
```
Timeout when executing method: alter_table_with_environment_context; 600525ms exceeds 600000ms
```

Cause Analysis

1. The MetaStore client connection times out. The default timeout interval for the connection between the MetaStore client and server is 600 seconds. On FusionInsight Manager, increase the value of **hive.metastore.client.socket.timeout** to **3600s**.
2. Another error is reported:
   ```
   Error: org.apache.hive.service.cli.HiveSQLException: Error while processing statement: FAILED:
   Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. Unable to alter table.
   java.net.SocketTimeoutException: Read timed out
   ```
   JDBC connection timeout interval of the MetaStore metadata. The default value is 60 ms.
3. Increase the value of **socketTimeout** in **javax.jdo.option.ConnectionURL** to **60000**. The initial error is still reported.
   ```
   Timeout when executing method: alter_table_with_environment_context;3600556ms exceeds
   3600000ms
   ```
4. Increase the values of parameters such as **hive.metastore.batch.retrieve.max**, **hive.metastore.batch.retrieve.table.partition.max**, and **dbservice.database.max.connections**. The problem persists.
5. It is suspected that the problem is caused by the GaussDB because adding a field will traverse each partition to execute **getPartitionColumnStatistics** and **alterPartition**.
6. Run the **gsql -p 20051 -U omm -W dbserverAdmin@123 -d hivemeta** command as user **omm** to log in to the Hive metabase.
7. Run **select * from pg_locks**. No lock wait is found.
8. Run **select * from pg_stat_activity;**. It is found that the process execution takes a long time.
   ```
   SELECT 'org.apache.hadoop.hive.metastore.model.MPartitionColumnStatistics'AS
   NUCLEUS_TYPE,A0.AVG_COL_LEN,A0."COLUMN_NAME",A0.COLUMN_TYPE,A0.DB_NAME,A0.BIG_DECI
   MAL_HIGH_VALUE,A0.BIG_DECIMAL_LOW_VALUE,A0.DOUBLE_HIGH_VALUE,A0.DOUBLE_LOW_VALUE,
   A0.LAST_ANALYZED,A0.LONG_HIGH_VALUE,A0.LONG_LOW_VALUE,A0.MAX_COL_LEN,A0.NUM_DISTIN
   CTS,A0.NUM_FALSES,A0.NUM_NULLS,A0.NUM_TRUES,A0.PARTITION_NAME,A0."TABLE_NAME",A0.CS_
   ID,A0.PARTITION_NAMEAS NUCORDER0 FROM PART_COL_STATS A0 WHERE A0."TABLE_NAME" =
   '$1' ANDA0.DB_NAME = '$2' AND A0.PARTITION_NAME = '$3' AND((((((A0."COLUMN_NAME" = '$4')
   OR (A0."COLUMN_NAME" ='$5')) OR (A0."COLUMN_NAME" = '$6')) OR (A0."COLUMN_NAME"
   ='$7')) OR (A0."COLUMN_NAME" = '$8')) OR (A0."COLUMN_NAME" ='$9')) ORDER BY NUCORDER0;
   ```

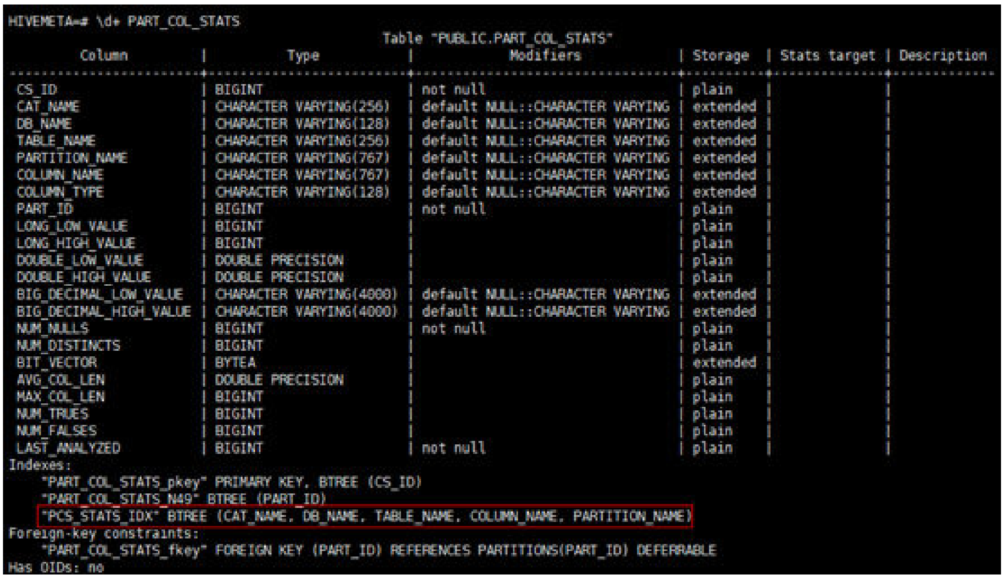


9. Run the **gs_guc reload -c log_min_duration_statement=100 -D /srv/BigData/dbdata_service/data/** command to start SQL recording. It is found that the execution duration of the **Run select * from pg_sta...** statement is **700 ms**, and more than 10,000 commands are executed because there are more than 10,000 partitions.
10. Add explain (analyze, verbose, timing, costs, buffers) before the SQL statement to analyze the execution plan. It is found that the entire table needs to be scanned during execution.



```
HIVEMETA=> explain (analyze,verbose,timing,costs,buffers) SELECT 'org.apache.hadoop.hive.metastore.model.MStorageDescriptor' AS NUCLEUS_TYPE,A0.INPUT_FORMAT,A0.IS_COMPRESSED,A0.IS_STOREDASSUBDIRECTO
RIES,A0.LOCATION,A0.NUM_BUCKETS,A0.OUTPUT_FORMAT,A0.SD_ID FROM SDS A0 WHERE A0.CD_ID = '463249' FETCH NEXT ROW ONLY;
                                                  QUERY PLAN
---------------------------------------------------------------------------------------------------------------------
 Limit  (cost=0.00..208.22 rows=1 width=218) (actual time=36.084..36.085 rows=1 loops=1)
   Output: ('org.apache.hadoop.hive.metastore.model.MStorageDescriptor'), INPUT_FORMAT, IS_COMPRESSED, IS_STOREDASSUBDIRECTORIES, LOCATION, NUM_BUCKETS, OUTPUT_FORMAT, SD_ID
   Buffers: shared hit=6720
   ->  Seq Scan on PUBLIC.SDS A0  (cost=0.00..52262.64 rows=251 width=218) (actual time=36.079..36.079 rows=1 loops=1)
         Output: 'org.apache.hadoop.hive.metastore.model.MStorageDescriptor', INPUT_FORMAT, IS_COMPRESSED, IS_STOREDASSUBDIRECTORIES, LOCATION, NUM_BUCKETS, OUTPUT_FORMAT, SD_ID
         Filter: (A0.CD_ID = 463249::BIGINT)
         Rows Removed by Filter: 154183
         Buffers: shared hit=6720
 Total runtime: 36.143 ms
(9 rows)
```

11. Check the index. It is found that the index does not meet the leftmost match rule.



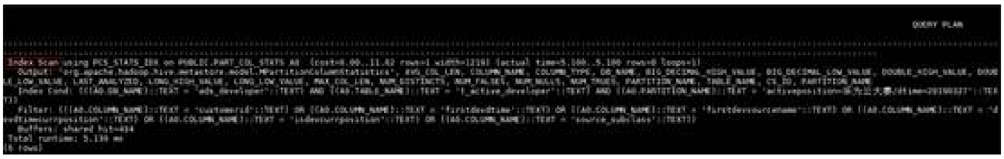
```
HIVEMETA=# \d+ PART_COL_STATS
                                              Table "PUBLIC.PART_COL_STATS"
         Column          |          Type           |               Modifiers               | Storage  | Stats target | Description
-------------------------+-------------------------+---------------------------------------+----------+--------------+-------------
 CS_ID                   | BIGINT                  | not null                              | plain    |              |
 CAT_NAME                | CHARACTER VARYING(256)  | default NULL::CHARACTER VARYING       | extended |              |
 DB_NAME                 | CHARACTER VARYING(128)  | default NULL::CHARACTER VARYING       | extended |              |
 TABLE_NAME              | CHARACTER VARYING(256)  | default NULL::CHARACTER VARYING       | extended |              |
 PARTITION_NAME          | CHARACTER VARYING(767)  | default NULL::CHARACTER VARYING       | extended |              |
 COLUMN_NAME             | CHARACTER VARYING(767)  | default NULL::CHARACTER VARYING       | extended |              |
 COLUMN_TYPE             | CHARACTER VARYING(128)  | default NULL::CHARACTER VARYING       | extended |              |
 PART_ID                 | BIGINT                  | not null                              | plain    |              |
 LONG_LOW_VALUE          | BIGINT                  |                                       | plain    |              |
 LONG_HIGH_VALUE         | BIGINT                  |                                       | plain    |              |
 DOUBLE_LOW_VALUE        | DOUBLE PRECISION        |                                       | plain    |              |
 DOUBLE_HIGH_VALUE       | DOUBLE PRECISION        |                                       | plain    |              |
 BIG_DECIMAL_LOW_VALUE   | CHARACTER VARYING(4000) | default NULL::CHARACTER VARYING       | extended |              |
 BIG_DECIMAL_HIGH_VALUE  | CHARACTER VARYING(4000) | default NULL::CHARACTER VARYING       | extended |              |
 NUM_NULLS               | BIGINT                  | not null                              | plain    |              |
 NUM_DISTINCTS           | BIGINT                  |                                       | plain    |              |
 BIT_VECTOR              | BYTEA                   |                                       | extended |              |
 AVG_COL_LEN             | DOUBLE PRECISION        |                                       | plain    |              |
 MAX_COL_LEN             | BIGINT                  |                                       | plain    |              |
 NUM_TRUES               | BIGINT                  |                                       | plain    |              |
 NUM_FALSES              | BIGINT                  |                                       | plain    |              |
 LAST_ANALYZED           | BIGINT                  | not null                              | plain    |              |
Indexes:
    "PART_COL_STATS_pkey" PRIMARY KEY, BTREE (CS_ID)
    "PART_COL_STATS_N49" BTREE (PART_ID)
    "PCS_STATS_IDX" BTREE (CAT_NAME, DB_NAME, TABLE_NAME, COLUMN_NAME, PARTITION_NAME)
Foreign-key constraints:
    "PART_COL_STATS_fkey" FOREIGN KEY (PART_ID) REFERENCES PARTITIONS(PART_ID) DEFERRABLE
Has OIDs: no
```

Procedure

1. Rebuild an index.

```
su – omm
gsql -p 20051 -U omm -W dbserverAdmin@123 -d hivemeta
DROP INDEX PCS_STATS_IDX;
CREATE INDEX PCS_STATS_IDX ON PART_COL_STATS(DB_NAME, TABLE_NAME, COLUMN_NAME,
PARTITION_NAME);
CREATE INDEX SDS_N50 ON SDS(CD_ID);
```

2. Check the execution plan again. It is found that the statement can be indexed and executed within 5 ms (the original execution time is 700 ms). Add fields to the Hive table again. The fields can be added successfully.



```
                                                                                                   QUERY PLAN
-------------------------------------------------------------------------------------------------------------------------------------
 Index Scan using PCS_STATS_IDX on PUBLIC.PART_COL_STATS A0  (cost=0.00..11.62 rows=1 width=1218) (actual time=5.100..5.100 rows=0 loops=1)
   Output: 'org.apache.hadoop.hive.metastore.model.MPartitionColumnStatistics', AVG_COL_LEN, COLUMN_NAME, COLUMN_TYPE, DB_NAME, BIG_DECIMAL_HIGH_VALUE, BIG_DECIMAL_LOW_VALUE, DOUBLE_HIGH_VALUE, DOUB
LE_LOW_VALUE, LAST_ANALYZED, LONG_HIGH_VALUE, LONG_LOW_VALUE, MAX_COL_LEN, NUM_DISTINCTS, NUM_FALSES, NUM_NULLS, NUM_TRUES, PARTITION_NAME, TABLE_NAME, CS_ID, PARTITION_NAME
   Index Cond: (((A0.DB_NAME)::TEXT = 'ads_developer'::TEXT) AND ((A0.TABLE_NAME)::TEXT = 't_active_developer'::TEXT) AND ((A0.PARTITION_NAME)::TEXT = 'activeposition=[illegible]/dtime=20190327'::TEX
T))
   Filter: ((((A0.COLUMN_NAME)::TEXT = 'customerid'::TEXT) OR ((A0.COLUMN_NAME)::TEXT = 'firstdevdtime'::TEXT) OR ((A0.COLUMN_NAME)::TEXT = 'firstdevsourcename'::TEXT) OR ((A0.COLUMN_NAME)::TEXT = 'd
evdtimesourposition'::TEXT) OR ((A0.COLUMN_NAME)::TEXT = 'indevsourposition'::TEXT) OR ((A0.COLUMN_NAME)::TEXT = 'source_subclass'::TEXT))
   Buffers: shared hit=434
 Total runtime: 5.138 ms
(6 rows)
```

15.10.43 Failed to Restart the Hive Service

Issue

After the Hive service configuration is modified, the configuration fails to be saved. The configuration status of the Hive service on Manager is **Failed**.

Symptom

User A opens the Hive configuration file in the background of the MRS node and does not close the file. User B modifies the Hive configuration item in **Service Management** on the MRS Manager page, saves the configuration, and restarts the Hive service. However, the configuration fails to be saved and the Hive service fails to be started.

Cause Analysis

When user B modifies the configuration on the MRS Manager page, the configuration file is opened by user A in the background of an MRS node. As a result, the configuration file cannot be replaced and the Hive service fails to be started.

Procedure

Step 1 Manually close the Hive configuration file opened in the background of the cluster node.

Step 2 Modify the Hive configuration on MRS Manager and save the configuration.

Step 3 Restart the Hive service.

----End

15.10.44 Hive Failed to Delete a Table

Issue

Hive fails to delete a table.

Symptom

Partitioning a Hive table by two columns may eventually generate over 20,000 partition files. As a result, the user fails to execute the **truncate table ${TableName}** or **drop table ${TableName}** statement to delete table data.

Cause Analysis

The file deletion operations are executed by a single thread serially. If the Hive partitioned tables have too many partition files, a large amount of metadata is stored in the metadata database. It takes a long time to delete metadata when a statement is executed to delete table data. As a result, the deletion cannot be complete within the specified timeout period, and the operation fails.

NOTE

You can log in to FusionInsight Manager and choose **Cluster** > **Services** > **Hive**. On the Hive page, choose **Configuration** > **All Configurations**, choose **ServerInit** under **MetaStore(Role)** in the navigation tree, and view the **hive.metastore.client.socket.timeout** parameter value in the right pane. This value is the timeout period. You can view the default value in the **Description** column.

Procedure

Step 1 (Optional, perform this step for an internal table) Use **alter table ${TableName} set TBLPROPERTIES('EXTERNAL'='true')** to convert it into an external table. In this way, only its metadata but not data stored on the HDFS is deleted, saving the table deletion time.

Step 2 (Optional, perform this step to use the same table name) Run the **show create table ${TableName}** command to export the table structure, and then run the **ALTER TABLE ${TableName} RENAME TO ${new_table_name};** command to rename the table. In this way, you can create a table that is the same as the original one.

Step 3 Run the **hdfs dfs –rm –r –f ${hdfs_path}** command to delete table data from the HDFS.

Step 4 Use **alter table ${Table_Name} drop partition (${PartitionName}<'XXXX', ${PartitionName}>'XXXX');** in Hive to delete partitions and reduce the number of files. The deletion conditions can be flexibly configured.

Step 5 When the number of rest partitions is smaller than 1,000, run the **drop table ${TableName}** command to delete the table.

----End

Summary and Suggestions

Hive partitioning can improve query efficiency. However, you should properly plan the partitioning policies to prevent a large number of small files from being generated.

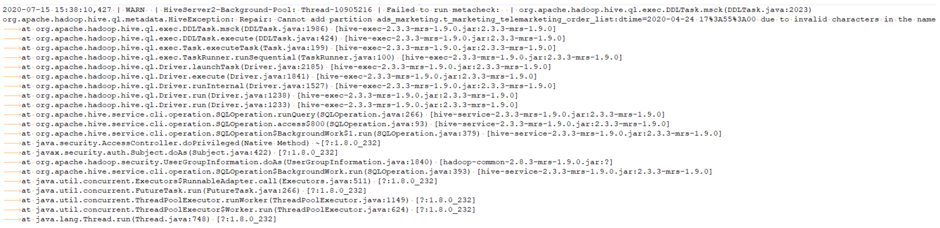
15.10.45 An Error Is Reported When msck repair table table_name Is Run on Hive

Symptom

When **msck repair table table_name** is run on Hive, the error message "FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask (state=08S01,code=1)" is displayed.

Possible Causes

A directory in the HiveServer log file **/var/log/Bigdata/hive/hiveserver/hive.log** does not comply with the partition format.



```
2020-07-15 15:38:10,427 | WARN  | HiveServer2-Background-Pool: Thread-10905216 | Failed to run metacheck:  | org.apache.hadoop.hive.ql.exec.DDLTask.msck(DDLTask.java:2023)
org.apache.hadoop.hive.ql.metadata.HiveException: Repair: Cannot add partition ads_marketing.t_marketing_telemarketing_order_list:dtime=2020-04-24 17%3A55%3A00 due to invalid characters in the name
    at org.apache.hadoop.hive.ql.exec.DDLTask.msck(DDLTask.java:1986) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
    at org.apache.hadoop.hive.ql.exec.DDLTask.execute(DDLTask.java:424) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
    at org.apache.hadoop.hive.ql.exec.Task.executeTask(Task.java:199) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
    at org.apache.hadoop.hive.ql.exec.TaskRunner.runSequential(TaskRunner.java:100) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
    at org.apache.hadoop.hive.ql.Driver.launchTask(Driver.java:2185) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
    at org.apache.hadoop.hive.ql.Driver.execute(Driver.java:1841) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
    at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1527) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
    at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1238) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
    at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1233) [hive-exec-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
    at org.apache.hive.service.cli.operation.SQLOperation.runQuery(SQLOperation.java:266) [hive-service-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
    at org.apache.hive.service.cli.operation.SQLOperation.access$800(SQLOperation.java:93) [hive-service-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
    at org.apache.hive.service.cli.operation.SQLOperation$BackgroundWork$1.run(SQLOperation.java:379) [hive-service-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
    at java.security.AccessController.doPrivileged(Native Method) ~[?:1.8.0_232]
    at javax.security.auth.Subject.doAs(Subject.java:422) [?:1.8.0_232]
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1840) [hadoop-common-2.8.3-mrs-1.9.0.jar:?]
    at org.apache.hive.service.cli.operation.SQLOperation$BackgroundWork.run(SQLOperation.java:393) [hive-service-2.3.3-mrs-1.9.0.jar:2.3.3-mrs-1.9.0]
    at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:511) [?:1.8.0_232]
    at java.util.concurrent.FutureTask.run(FutureTask.java:266) [?:1.8.0_232]
    at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149) [?:1.8.0_232]
    at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624) [?:1.8.0_232]
    at java.lang.Thread.run(Thread.java:748) [?:1.8.0_232]
```

Procedure

- Method 1: Delete the incorrect file or directory.
- Method 2: Run the **set hive.msck.path.validation=skip** command to skip invalid directories.

15.10.46 How Do I Release Disk Space After Dropping a Table in Hive?

Issue

After a user runs the **drop** command on the Hive CLI to drop a table and then uses the **hdfs dfsadmin -report** command to check the disk space, the command output shows that the table is not deleted.

Cause Analysis

The **drop** command executed on the Hive CLI deletes only the table structure of the external table, but not the table data stored in HDFS.

Procedure

Step 1 Log in to the node where the client is installed as user **root** and authenticate the component user.

cd *Client installation directory*

source bigdata_env

kinit *Component service user* (Skip this step for clusters with Kerberos authentication disabled.)

Step 2 Run the following command to delete the table stored in HDFS:

hadoop fs -rm hdfs://hacluster/*Path of the table*

----End

15.10.47 Connection Timeout During SQL Statement Execution on the Client

Symptom

The SQL statement fails to be executed on the client, and the error message "Timed out waiting for a free available connection" is displayed.

Possible Causes

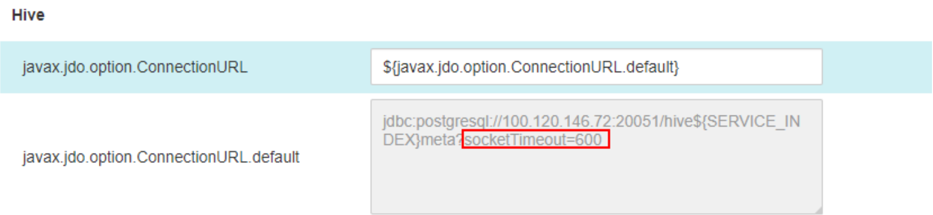
A large number of DBService connections exist, and obtaining connections times out.

Procedure

Step 1 Check whether the client uses the Spark-SQL client to execute SQL statements.

- If yes, check the timeout parameter in the URL, change the value to **600**, and go to **Step 7**.
- If the alarm is not cleared, go to **Step 2**.

Step 2 Log in to FusionInsight Manager, choose **Cluster** > **Services** > **Hive** > **Configuration** > **All Configurations**, search for **javax.jdo.option.ConnectionURL**, and check whether the value of the timeout parameter is less than **600**.



Hive

javax.jdo.option.ConnectionURL	${javax.jdo.option.ConnectionURL.default}
javax.jdo.option.ConnectionURL.default	jdbc:postgresql://100.120.146.72:20051/hive${SERVICE_INDEX}meta?socketTimeout=600

NOTE

This parameter exists in Hive, HiveServer, MetaStore, and WebHCat. Ensure that the parameter values are the same.

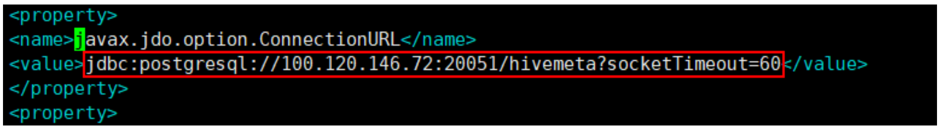
- If yes, go to **Step 3**.
- If no, go to **Step 7**.

Step 3 Check whether the value of **javax.jdo.option.ConnectionURL** is **${javax.jdo.option.ConnectionURL.default}**.

- If yes, go to **Step 4**.
- If no, change the timeout parameter in the URL to **600**, click **Save**, and go to **Step 7**.

Step 4 Click **Instance**, select any HiveServer instance, and log in to the instance node as user **root**.

Step 5 Open the **${BIGDATA_HOME}/FusionInsight_Current/*HiveServer/etc/hivemetastore-site.xml** configuration file, find the **javax.jdo.option.ConnectionURL** parameter, and copy its value.



```
<property>
<name>javax.jdo.option.ConnectionURL</name>
<value>jdbc:postgresql://100.120.146.72:20051/hivemeta?socketTimeout=60</value>
</property>
<property>
```

Step 6 Log in to FusionInsight Manager, choose **Cluster** > **Services** > **Hive** > **Configuration** > **All Configurations**, search for **javax.jdo.option.ConnectionURL**, change its value to the URL copied in **Step 5**, change the timeout parameter to **600**, and click **Save**.

NOTE

This parameter exists in Hive, HiveServer, MetaStore, and WebHCat. Ensure that the parameter values are the same.

Step 7 Choose **Cluster** > **Services** > **Hive** > **Configuration** > **All Configurations**, search for **maxConnectionsPerPartition**, and check whether its value is less than **100**.

- If yes, change the value to **100**, click **Save**, and go to **Step 8**.
- If no, go to **Step 8**.

Step 8 If parameters are modified in the preceding steps, choose **Cluster** > **Services** > **Hive** > **Dashboard** and choose **More** > **Service Rolling Restart**. If the parameters are not modified, skip this step.

----End

15.10.48 WebHCat Failed to Start Due to Abnormal Health Status

Issue

The WebHCat instance fails to be started.

Symptom

On Manager, the health status of the WebHCat instance is **Faulty**, and alarm **ALM-12007 Process Fault** is generated for the WebHCat instance of the Hive service. An error is reported when the Hive service is restarted.

Error messages "Service not found in Kerberos database" and "Address already in use" are contained in the **/var/log/Bigdata/hive/webhcat/webhcat.log** file of the WebHCat instance.

Procedure

Step 1 Log in to each node where the WebHCat instance resides and check whether the mapping between IP addresses and hostnames in the **/etc/hosts** file is correct. The WebHCat configurations in the **/etc/hostname** and **/etc/HOSTNAME** files must be the same as those in the **/etc/hosts** file. If they are different, manually modify them.

NOTE

To view the mapping between the IP addresses and hostnames of the WebHCat instance, log in to FusionInsight Manager and choose **Cluster** > **Services** > **Hive** > **Instance**.

Step 2 Log in to any node where the WebHCat instance resides and run the following command to switch to user **omm**:

su - omm

Step 3 Run the following command to check whether the WebHCat process exists:

ps -ef|grep webhcat|grep -v grep

If it does, run the following command to kill it:

kill -9 ${webhcat_pid}

Step 4 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **Hive** . On the page that is displayed, click the **Instance** tab. The select all WebHCat instances, click **More**, and select **Restart Instance**. Wait until WebHCat is restarted successfully.

----End

15.10.49 WebHCat Failed to Start Because the mapred-default.xml File Cannot Be Parsed

Issue

The Hive service of MRS is faulty. After the Hive service is restarted, the HiveServer and WebHCat processes on the Master2 node fail to start, but the processes on the Master1 node are normal.

Cause Analysis

Log in to the Master2 node and check the **/var/log/Bigdata/hive/hiveserver/hive.log** file. It is found that HiveServer keeps loading **/opt/Bigdata/*/*_HiveServer/etc/hive-site.xml**. Check the **/var/log/Bigdata/hive/hiveserver/hiveserver.out** log generated when HiveServer exits. It is found that an exception occurs when the **mapred-default.xml** file is parsed.

Procedure

Step 1 Log in to the Master2 node and run the following command to query the path of **mapred-default.xml**:

find /opt/ -name 'mapred-default.xml'

The configuration file is in the **/opt/Bigdata/*/*_WebHCat/etc/** directory but is empty.

Step 2 Log in to the Master1 node, copy the **/opt/Bigdata/*/*_WebHCat/etc/mapred-default.xml** file to the Master2 node, and change the owner group of the file to **omm:wheel**.

Step 3 Log in to Manager and restart the abnormal HiveServer and WebHCat instances.

----End

15.11 Using Hue

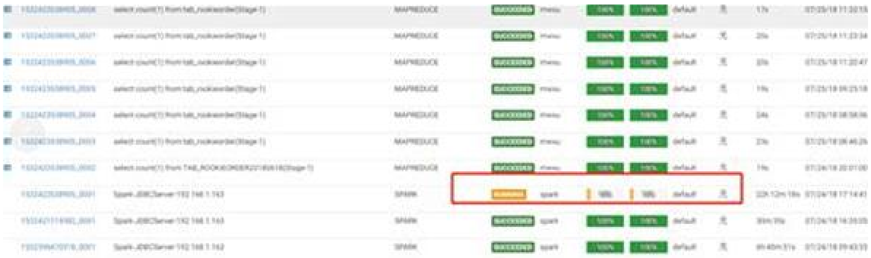
15.11.1 A Job Is Running on Hue

Issue

The customer finds that a job is running on Hue.

Symptom

After the customer's MRS is installed, the job is running on Hue but the running job is not operated by the customer.



Cause Analysis

This job is a permanent job generated when the system connects to JDBC after Spark is started.

Procedure

This is not a problem. No handling is required.

15.11.2 HQL Fails to Be Executed on Hue Using Internet Explorer

Symptom

Using Internet Explorer to access Hive Editor and execute all HQL statements on Hue fails and the system prompts "There was an error with your query".

Cause Analysis

Internet Explorer has functional problems and cannot process AJAX POST requests containing form data in 307 redirection. Use a compatible browser.

Solution

Use Google Chrome 21 or later.

15.11.3 Hue (Active) Cannot Open Web Pages

Symptom

The following information is displayed on the web UI of Hue (active):

```
Service Unavailable
The server is temporarily unable to service your request due to maintenance downtime or capacity
problems. Please try again later.
```

Cause Analysis

- The Hue configuration has expired.

- The configuration of the Hue service needs to be modified manually in a single-master cluster.

Solution

- If the Hue configuration has expired, restart the Hue service.
- Manually modify the Hue service configuration for a single-master cluster.
  a. Log in to the Master node.
  b. Run the **hostname -i** command to obtain the IP address of the local host.
  c. Run the following command to obtain the value of **HUE_FLOAT_IP**:

     grep "HUE_FLOAT_IP" ${BIGDATA_HOME}/*MRS*_Current/1_*/etc*/ENV_VARS,

     Replace *MRS* with the actual file name.
  d. Check whether the local IP address is the same as the value of **HUE_FLOAT_IP**. If they are different, change the value of **HUE_FLOAT_IP** to the local IP address.
  e. Restart the Hue service.

15.11.4 Failed to Access the Hue Web UI

Issue

An error page is displayed when the Hue web UI is accessed.

Symptom

The following error information is displayed on the Hue web UI:

```
503 Service Unavailable
The server is temporarily unable to service your requster due to maintenance downtime or capacity
problems.Please try again later.
```

Cause Analysis

- The Hue configuration has expired.
- The configuration of the Hue service needs to be modified manually in a single-master cluster.

Procedure

Step 1 Log in to the Master node.

Step 2 Run the **hostname -i** command to obtain the IP address of the local host.

Step 3 Run the following command to obtain the value of **HUE_FLOAT_IP**:

grep "HUE_FLOAT_IP" ${BIGDATA_HOME}/*MRS*_Current/1_*/etc*/ENV_VARS, where *MRS* is subject to the actual file name.

Step 4 Check whether the local IP address is the same as the value of **HUE_FLOAT_IP**. If they are different, change the value of **HUE_FLOAT_IP** to the local IP address.

Step 5 Restart the Hue service.

----End

15.11.5 HBase Tables Cannot Be Loaded on the Hue Web UI

Issue

After Hive data is imported to HBase on the Hue page, an error message is displayed, indicating that the HBase table cannot be detected.

Symptom

In the Kerberos cluster, the IAM sub-account does not have sufficient permissions. As a result, the HBase table cannot be loaded.

Cause Analysis

The IAM subaccount does not have sufficient permissions.

Procedure

MRS Manager:

Step 1 Log in to MRS Manager.

Step 2 Choose **System** > **Manage User**.

Step 3 Locate the row that contains the target user, and click **Modify**.

Step 4 Add the user to the **supergroup** group.

Step 5 Click **OK**. The modification is complete.

----End

FusionInsight Manager:

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 Locate the row that contains the target user, and click **Modify**.

Step 4 Add the user to the **supergroup** group.

Step 5 Click **OK**. The modification is complete.

----End

Summary and Suggestions

If Kerberos authentication is enabled for a cluster, "No data available" is displayed on the page. In this case, check the permission first.

15.12 Using Impala

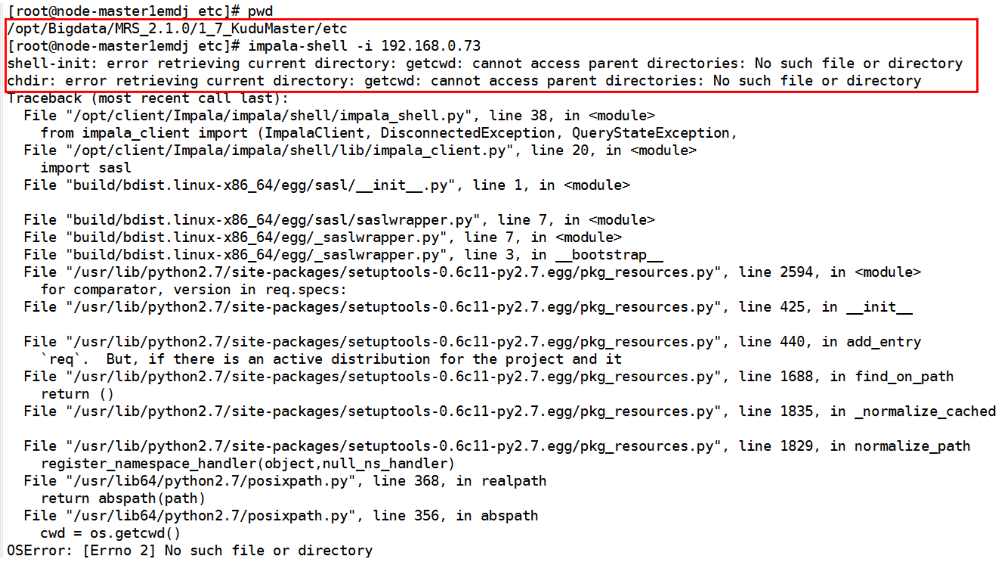
15.12.1 Failed to Connect to impala-shell

Issue

A user fails to connect to impala-shell.

Symptom

After a user modifies the configuration of any component on the component management page and restarts the service, the connection to impala-shell fails, and the error message "no such file/directory" is displayed.



```
[root@node-master1emdj etc]# pwd
/opt/Bigdata/MRS_2.1.0/1_7_KuduMaster/etc
[root@node-master1emdj etc]# impala-shell -i 192.168.0.73
shell-init: error retrieving current directory: getcwd: cannot access parent directories: No such file or directory
chdir: error retrieving current directory: getcwd: cannot access parent directories: No such file or directory
Traceback (most recent call last):
  File "/opt/client/Impala/impala/shell/impala_shell.py", line 38, in <module>
    from impala_client import (ImpalaClient, DisconnectedException, QueryStateException,
  File "/opt/client/Impala/impala/shell/lib/impala_client.py", line 20, in <module>
    import sasl
  File "build/bdist.linux-x86_64/egg/sasl/__init__.py", line 1, in <module>

  File "build/bdist.linux-x86_64/egg/sasl/saslwrapper.py", line 7, in <module>
  File "build/bdist.linux-x86_64/egg/_saslwrapper.py", line 7, in <module>
  File "build/bdist.linux-x86_64/egg/_saslwrapper.py", line 3, in __bootstrap__
  File "/usr/lib/python2.7/site-packages/setuptools-0.6c11-py2.7.egg/pkg_resources.py", line 2594, in <module>
    for comparator, version in req.specs:
  File "/usr/lib/python2.7/site-packages/setuptools-0.6c11-py2.7.egg/pkg_resources.py", line 425, in __init__

  File "/usr/lib/python2.7/site-packages/setuptools-0.6c11-py2.7.egg/pkg_resources.py", line 440, in add_entry
    `req`.  But, if there is an active distribution for the project and it
  File "/usr/lib/python2.7/site-packages/setuptools-0.6c11-py2.7.egg/pkg_resources.py", line 1688, in find_on_path
    return ()
  File "/usr/lib/python2.7/site-packages/setuptools-0.6c11-py2.7.egg/pkg_resources.py", line 1835, in _normalize_cached

  File "/usr/lib/python2.7/site-packages/setuptools-0.6c11-py2.7.egg/pkg_resources.py", line 1829, in normalize_path
    register_namespace_handler(object,null_ns_handler)
  File "/usr/lib64/python2.7/posixpath.py", line 368, in realpath
    return abspath(path)
  File "/usr/lib64/python2.7/posixpath.py", line 356, in abspath
    cwd = os.getcwd()
OSError: [Errno 2] No such file or directory
```

Cause Analysis

After the service configuration is modified and the service is restarted, some directory structures of the service, such as the etc directory, are deleted and recreated. If the directory is etc or its subdirectory before the service is restarted, some system variables or parameters cannot be found when impala-shell is executed in the original directory because the directory is recreated after the service is restarted. As a result, impala-shell fails to be connected.

Procedure

Switch to any existing directory and reconnect to impala-shell.

15.12.2 Failed to Create a Kudu Table

Issue

An error occurs when a user creates a Kudu table.

Symptom

When a user creates table in a new cluster, the following error message is displayed: [Cloudera]ImpalaJDBCDriver ERROR processing query/statement. Error

Code: 0, SQL state: TStatus(statusCode:ERROR_STATUS, sqlState:HY000, errorMessage:AnalysisException: Table property 'kudu.master_addresses' is required when the impalad startup flag -kudu_master_hosts is not used.

Cause Analysis

The user does not specify **kudu.master_addresses** in the Impala SQL statement.

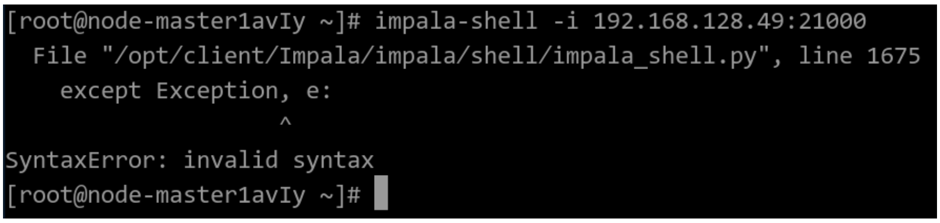
Procedure

Specify **kudu.master_addresses** when creating a Kudu table.

15.12.3 Failed to Log In to the Impala Client

Issue

Error information similar to the following is displayed when a user runs the Impala client.



```
[root@node-master1avIy ~]# impala-shell -i 192.168.128.49:21000
  File "/opt/client/Impala/impala/shell/impala_shell.py", line 1675
    except Exception, e:
                    ^
SyntaxError: invalid syntax
[root@node-master1avIy ~]# 
```

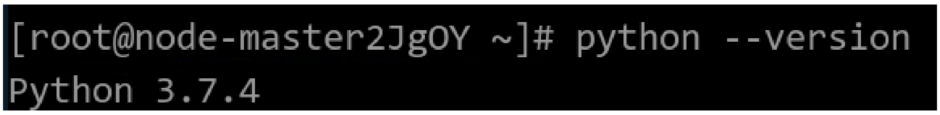
Cause Analysis

The latest MRS cluster uses EulerOS 2.9 or later, which provides only Python 3. However, the Impala client is implemented based on Python 2 and is incompatible with some syntax of Python 3. As a result, an error occurs when the Impala client is running. You can manually install Python 2 to solve this problem.

Procedure

Step 1 Log in to the Impala node as user **root** and run the following command to check its Python version:

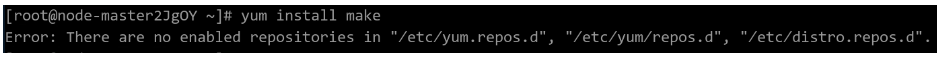
python --version



```
[root@node-master2JgOY ~]# python --version
Python 3.7.4
```

Step 2 Run the **yum install make** command to check whether yum is available.

- If the following error is reported, the yum configuration is incorrect. Go to **Step 3**.

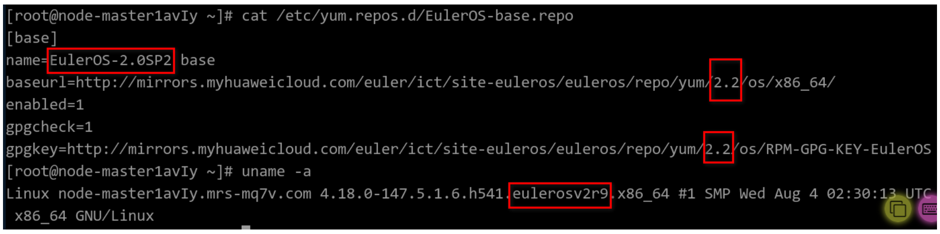


```
[root@node-master2JgOY ~]# yum install make
Error: There are no enabled repositories in "/etc/yum.repos.d", "/etc/yum/repos.d", "/etc/distro.repos.d".
```

- If no error is reported, go to **Step 4**.

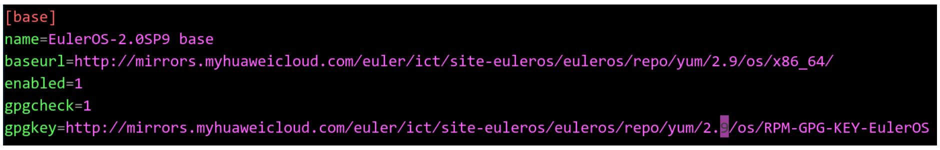
Step 3 Run the **cat /etc/yum.repos.d/EulerOS-base.repo** command to check whether the yum source matches the system version. If they do not match, modify them.

Before modification



```
[root@node-master1avIy ~]# cat /etc/yum.repos.d/EulerOS-base.repo
[base]
name=EulerOS-2.0SP2 base
baseurl=http://mirrors.myhuaweicloud.com/euler/ict/site-euleros/euleros/repo/yum/2.2/os/x86_64/
enabled=1
gpgcheck=1
gpgkey=http://mirrors.myhuaweicloud.com/euler/ict/site-euleros/euleros/repo/yum/2.2/os/RPM-GPG-KEY-EulerOS
[root@node-master1avIy ~]# uname -a
Linux node-master1avIy.mrs-mq7v.com 4.18.0-147.5.1.6.h541.eulerosv2r9.x86_64 #1 SMP Wed Aug 4 02:30:13 UTC
 x86_64 GNU/Linux
```

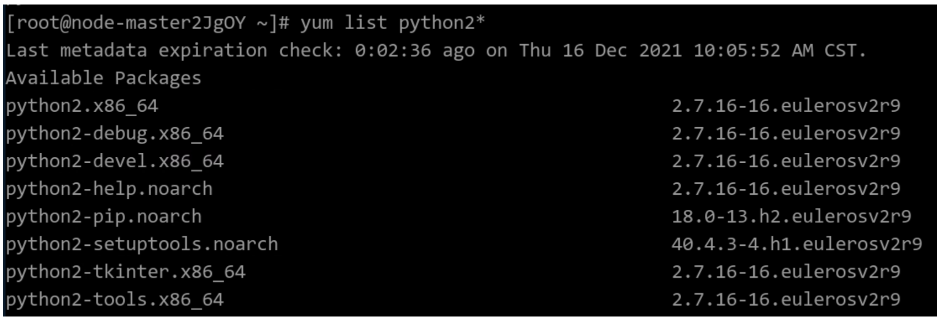
After modification



```
[base]
name=EulerOS-2.0SP9 base
baseurl=http://mirrors.myhuaweicloud.com/euler/ict/site-euleros/euleros/repo/yum/2.9/os/x86_64/
enabled=1
gpgcheck=1
gpgkey=http://mirrors.myhuaweicloud.com/euler/ict/site-euleros/euleros/repo/yum/2.9/os/RPM-GPG-KEY-EulerOS
```

Step 4 Run the following command to check for the software whose name starts with **python2** in the yum source:

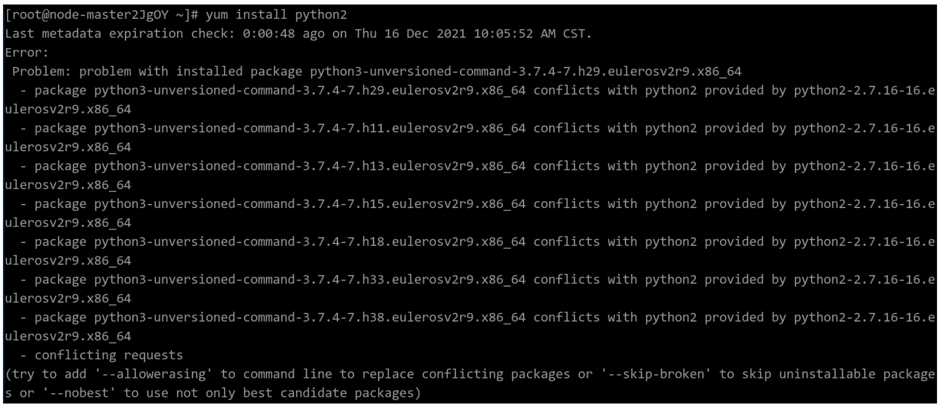
yum list python2*



```
[root@node-master2JgOY ~]# yum list python2*
Last metadata expiration check: 0:02:36 ago on Thu 16 Dec 2021 10:05:52 AM CST.
Available Packages
python2.x86_64                                  2.7.16-16.eulerosv2r9
python2-debug.x86_64                            2.7.16-16.eulerosv2r9
python2-devel.x86_64                            2.7.16-16.eulerosv2r9
python2-help.noarch                             2.7.16-16.eulerosv2r9
python2-pip.noarch                              18.0-13.h2.eulerosv2r9
python2-setuptools.noarch                       40.4.3-4.h1.eulerosv2r9
python2-tkinter.x86_64                          2.7.16-16.eulerosv2r9
python2-tools.x86_64                            2.7.16-16.eulerosv2r9
```

Step 5 Run the following command to install Python 2:

yum install python2

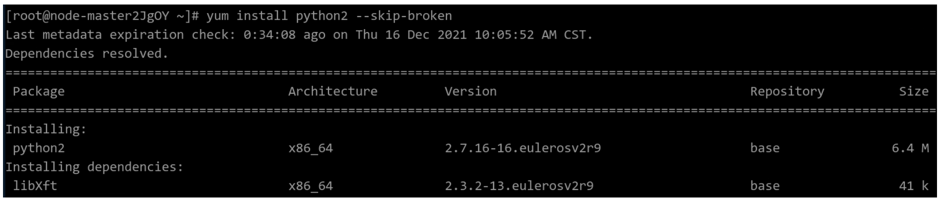


```
[root@node-master2JgOY ~]# yum install python2
Last metadata expiration check: 0:00:48 ago on Thu 16 Dec 2021 10:05:52 AM CST.
Error:
 Problem: problem with installed package python3-unversioned-command-3.7.4-7.h29.eulerosv2r9.x86_64
  - package python3-unversioned-command-3.7.4-7.h29.eulerosv2r9.x86_64 conflicts with python2 provided by python2-2.7.16-16.e
ulerosv2r9.x86_64
  - package python3-unversioned-command-3.7.4-7.h11.eulerosv2r9.x86_64 conflicts with python2 provided by python2-2.7.16-16.e
ulerosv2r9.x86_64
  - package python3-unversioned-command-3.7.4-7.h13.eulerosv2r9.x86_64 conflicts with python2 provided by python2-2.7.16-16.e
ulerosv2r9.x86_64
  - package python3-unversioned-command-3.7.4-7.h15.eulerosv2r9.x86_64 conflicts with python2 provided by python2-2.7.16-16.e
ulerosv2r9.x86_64
  - package python3-unversioned-command-3.7.4-7.h18.eulerosv2r9.x86_64 conflicts with python2 provided by python2-2.7.16-16.e
ulerosv2r9.x86_64
  - package python3-unversioned-command-3.7.4-7.h33.eulerosv2r9.x86_64 conflicts with python2 provided by python2-2.7.16-16.e
ulerosv2r9.x86_64
  - package python3-unversioned-command-3.7.4-7.h38.eulerosv2r9.x86_64 conflicts with python2 provided by python2-2.7.16-16.e
ulerosv2r9.x86_64
  - conflicting requests
(try to add '--allowerasing' to command line to replace conflicting packages or '--skip-broken' to skip uninstallable package
s or '--nobest' to use not only best candidate packages)
```

Python 3 has been installed in the current system. If you directly install Python 2, a conflict message is displayed.

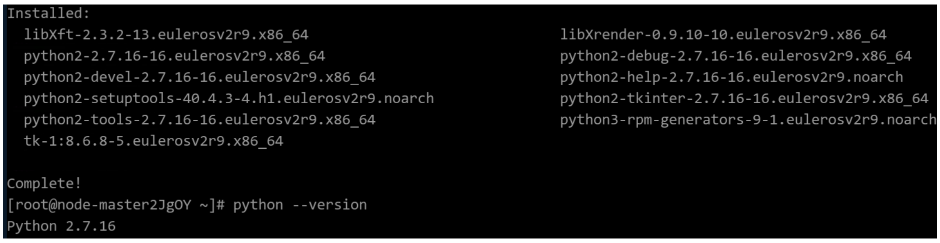
You can select **--allowerasing** or **--skip-broken** for the installation. For example:

yum install python2 --skip-broken



```
[root@node-master2JgOY ~]# yum install python2 --skip-broken
Last metadata expiration check: 0:34:08 ago on Thu 16 Dec 2021 10:05:52 AM CST.
Dependencies resolved.
=====================================================================================================================
 Package                            Architecture        Version                               Repository        Size
=====================================================================================================================
Installing:
 python2                            x86_64              2.7.16-16.eulerosv2r9                 base             6.4 M
Installing dependencies:
 libXft                             x86_64              2.3.2-13.eulerosv2r9                  base              41 k
```

After the installation is complete, the Python version is automatically changed to python2, as shown in the following figure.

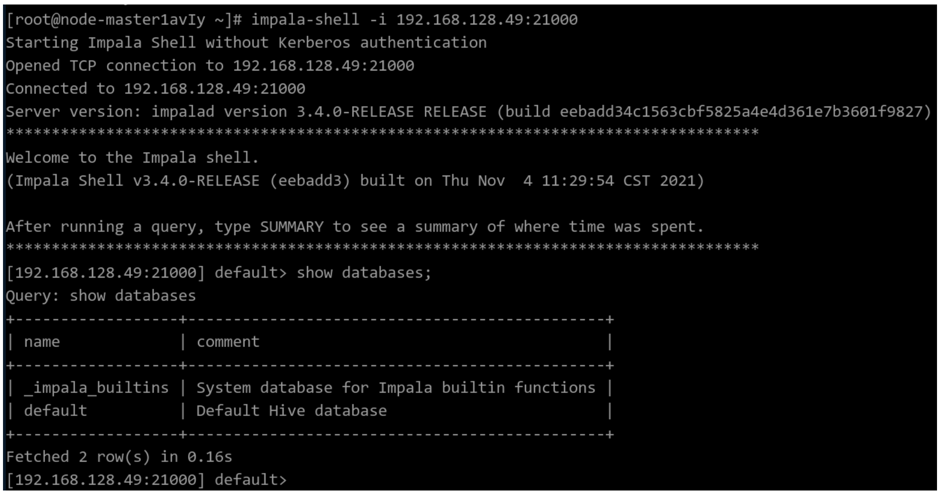


```
Installed:
  libXft-2.3.2-13.eulerosv2r9.x86_64                              libXrender-0.9.10-10.eulerosv2r9.x86_64
  python2-2.7.16-16.eulerosv2r9.x86_64                            python2-debug-2.7.16-16.eulerosv2r9.x86_64
  python2-devel-2.7.16-16.eulerosv2r9.x86_64                      python2-help-2.7.16-16.eulerosv2r9.noarch
  python2-setuptools-40.4.3-4.h1.eulerosv2r9.noarch               python2-tkinter-2.7.16-16.eulerosv2r9.x86_64
  python2-tools-2.7.16-16.eulerosv2r9.x86_64                      python3-rpm-generators-9-1.eulerosv2r9.noarch
  tk-1:8.6.8-5.eulerosv2r9.x86_64

Complete!
[root@node-master2JgOY ~]# python --version
Python 2.7.16
```

If Python 2 is installed successfully but the displayed Python version is incorrect, run the following command to create the **/usr/bin/python** soft link for **/usr/bin/python2**:

ln -s /usr/bin/python2 /usr/bin/python

Step 6 Verify that the Impala client is available.



```
[root@node-master1avIy ~]# impala-shell -i 192.168.128.49:21000
Starting Impala Shell without Kerberos authentication
Opened TCP connection to 192.168.128.49:21000
Connected to 192.168.128.49:21000
Server version: impalad version 3.4.0-RELEASE RELEASE (build eebadd34c1563cbf5825a4e4d361e7b3601f9827)
***********************************************************************************
Welcome to the Impala shell.
(Impala Shell v3.4.0-RELEASE (eebadd3) built on Thu Nov  4 11:29:54 CST 2021)

After running a query, type SUMMARY to see a summary of where time was spent.
***********************************************************************************
[192.168.128.49:21000] default> show databases;
Query: show databases
+------------------+----------------------------------------------+
| name             | comment                                      |
+------------------+----------------------------------------------+
| _impala_builtins | System database for Impala builtin functions |
| default          | Default Hive database                        |
+------------------+----------------------------------------------+
Fetched 2 row(s) in 0.16s
[192.168.128.49:21000] default>
```

----End

15.13 Using Kafka

15.13.1 An Error Is Reported When Kafka Is Run to Obtain a Topic

Issue

An Error is reported when Kafka is run to obtain a topic.

Symptom

An error is reported when the Kafka is run to obtain topics. The error information is as follows:

```
ERROR org.apache.kafka.common.errors.InvalidReplicationFactorException: Replication factor: 2 larger than
available brokers: 0.
```

Possible Cause

The variable for obtaining the ZooKeeper address is incorrect due to special characters.

Procedure

Step 1 Log in to any Master node.

Step 2 Run the **cat /opt/client/Kafka/kafka/config/server.properties |grep '^zookeeper.connect ='** command to check the variable of the Zookeeper address.

Step 3 Run Kafka again to obtain the topic. Do not add any character to the variables obtained in **Step 2**.

----End

15.13.2 Flume Normally Connects to Kafka But Fails to Send Messages

Symptom

An MRS cluster is installed, and ZooKeeper, Flume, and Kafka are installed in the cluster.

Flume fails to send data to Kafka.

Possible Causes

1. The Kafka service is abnormal.
2. The IP address for Flume to connect to Kafka is incorrect.
3. The size of the message sent from Flume to Kafka exceeds the upper limit.

Cause Analysis

The possible reasons why Flume fails to send data to Kafka may be related to Flume or Kafka.

1. Check the Kafka service status and monitoring metrics on Manager.
   - MRS Manager: Log in to MRS Manager and choose **Services** > **Kafka**. Check the Kafka status. The status is **Good**, and the monitoring metrics are correctly displayed.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Kafka**.

     Check the Kafka status. It is found that the status is good and the monitoring metrics are correctly displayed.

2. Check the Flume log. The log contains **MessageSizeTooLargeException** information, as shown in the following:

```
2016-02-26 14:55:19,126 | WARN  | [SinkRunner-PollingRunner-DefaultSinkProcessor] |  Produce
request with correlation id 349829 failed due to [LOG,7]:
kafka.common.MessageSizeTooLargeException  | kafka.utils.Logging$class.warn(Logging.scala:83)
```

   The exception shows that the size of data written to Kafka by Flume exceeds the maximum message size specified by Kafka.

3. Check the maximum message size specified by Kafka on Manager.
   - MRS Manager portal: Log in to MRS Manager and choose **Services** > **Kafka** > **Configuration**.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Kafka** > **Configuration**.

   On the page that is displayed, set **Type** to **All**. All Kafka configurations are displayed. Enter **message.max.bytes** in the **Search** text box to search.

   In MRS, the maximum size of a message that can be received by the Kafka server is 1000012 bytes = 977 KB by default.

Solution

After confirmation with the customer, data sent by Flume contains messages over 1 MB. Adjust parameters on Kafka to enable the messages to be written to Kafka.

Step 1 Set **message.max.bytes** to a value that is larger than the current maximum size of the message to be written so that Kafka can receive all messages.

Step 2 Set **replica.fetch.max.bytes** to a value that is equal to or larger than the value of **message.max.bytes** so that replicas of partitions on different Brokers can be synchronized to all messages.

- MRS Manager portal: Log in to MRS Manager and choose **Services** > **Kafka** > **Configuration**.
- FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Kafka** > **Configuration**.

On the page that is displayed, set **Type** to **All**. All Kafka configurations are displayed. Enter **replica.fetch.max.bytes** in the **Search** text box to search.

Step 3 Click **Save** and restart the Kafka service to make Kafka configurations take effect.

Step 4 Set **fetch.message.max.bytes** to a value that is equal to or larger than the value of **message.max.bytes** for Consumer service applications to ensure that Consumers can consume all messages.

----End

15.13.3 Producer Failed to Send Data and Threw "NullPointerException"

Symptom

An MRS cluster has ZooKeeper and Kafka installed.

When the Producer client sends data to Kafka, it fails and throws "NullPointerException".

Possible Causes

1. The Kafka service is abnormal.
2. The **jass** and **keytab** files configured on the Producer client are incorrect.

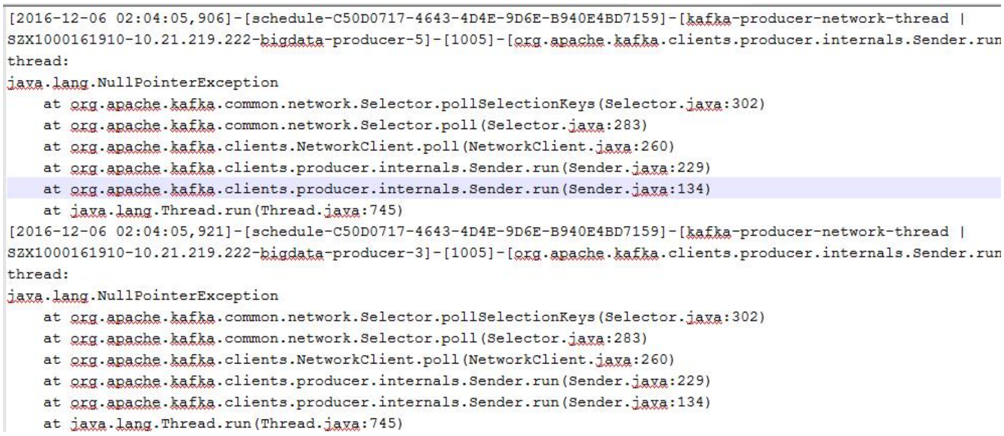
Cause Analysis

The possible causes may be related to Producer or Kafka.

1. Check the Kafka service status and monitoring metrics on Manager.
   - MRS Manager: Log in to MRS Manager and choose **Services** > **Kafka**. Check the Kafka status. The status is **Good**, and the monitoring metrics are correctly displayed.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the target cluster and choose **Service** > **Kafka**.

     Check the Kafka status. The status is good, and the monitoring metrics are correctly displayed.
2. Check the Producer client log. The log contains "NullPointerException", as shown in **Figure 15-42**.

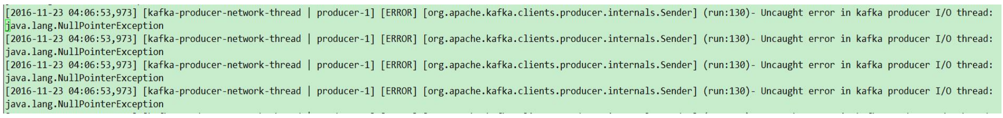
   Figure 15-42 Producer client log

   ```
   [2016-12-06 02:04:05,906]-[schedule-C50D0717-4643-4D4E-9D6E-B940E4BD7159]-[kafka-producer-network-thread |
   SZX1000161910-10.21.219.222-bigdata-producer-5]-[1005]-[org.apache.kafka.clients.producer.internals.Sender.run
   thread:
   java.lang.NullPointerException
       at org.apache.kafka.common.network.Selector.pollSelectionKeys(Selector.java:302)
       at org.apache.kafka.common.network.Selector.poll(Selector.java:283)
       at org.apache.kafka.clients.NetworkClient.poll(NetworkClient.java:260)
       at org.apache.kafka.clients.producer.internals.Sender.run(Sender.java:229)
       at org.apache.kafka.clients.producer.internals.Sender.run(Sender.java:134)
       at java.lang.Thread.run(Thread.java:745)
   [2016-12-06 02:04:05,921]-[schedule-C50D0717-4643-4D4E-9D6E-B940E4BD7159]-[kafka-producer-network-thread |
   SZX1000161910-10.21.219.222-bigdata-producer-3]-[1005]-[org.apache.kafka.clients.producer.internals.Sender.run
   thread:
   java.lang.NullPointerException
       at org.apache.kafka.common.network.Selector.pollSelectionKeys(Selector.java:302)
       at org.apache.kafka.common.network.Selector.poll(Selector.java:283)
       at org.apache.kafka.clients.NetworkClient.poll(NetworkClient.java:260)
       at org.apache.kafka.clients.producer.internals.Sender.run(Sender.java:229)
       at org.apache.kafka.clients.producer.internals.Sender.run(Sender.java:134)
       at java.lang.Thread.run(Thread.java:745)
   ```

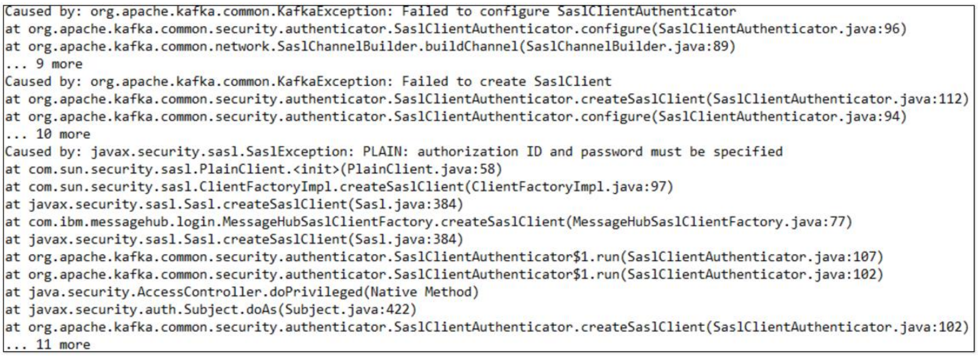
   Alternatively, the log contains only "NullPointerException" but no stack information. The problem is caused by JDK self-protection. If much information is printed for the same stack, the JDK self-protection is triggered and stack information is no longer printed, as shown in **Figure 15-43**.

   Figure 15-43 Error information

   ```
   [2016-11-23 04:06:53,973] [kafka-producer-network-thread | producer-1] [ERROR] [org.apache.kafka.clients.producer.internals.Sender] (run:130)- Uncaught error in kafka producer I/O thread:
   java.lang.NullPointerException
   [2016-11-23 04:06:53,973] [kafka-producer-network-thread | producer-1] [ERROR] [org.apache.kafka.clients.producer.internals.Sender] (run:130)- Uncaught error in kafka producer I/O thread:
   java.lang.NullPointerException
   [2016-11-23 04:06:53,973] [kafka-producer-network-thread | producer-1] [ERROR] [org.apache.kafka.clients.producer.internals.Sender] (run:130)- Uncaught error in kafka producer I/O thread:
   java.lang.NullPointerException
   [2016-11-23 04:06:53,973] [kafka-producer-network-thread | producer-1] [ERROR] [org.apache.kafka.clients.producer.internals.Sender] (run:130)- Uncaught error in kafka producer I/O thread:
   java.lang.NullPointerException
   ```
3. Check the Producer client log. Error information "Failed to configure SaslClientAuthenticator" is displayed, as shown in **Figure 15-44**.

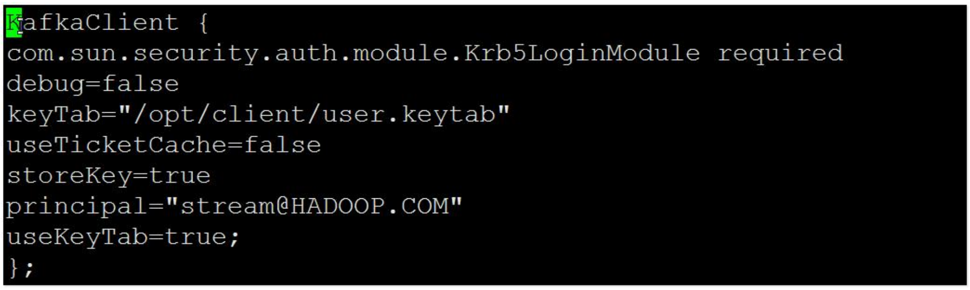
Figure 15-44 Error log



```
Caused by: org.apache.kafka.common.KafkaException: Failed to configure SaslClientAuthenticator
at org.apache.kafka.common.security.authenticator.SaslClientAuthenticator.configure(SaslClientAuthenticator.java:96)
at org.apache.kafka.common.network.SaslChannelBuilder.buildChannel(SaslChannelBuilder.java:89)
... 9 more
Caused by: org.apache.kafka.common.KafkaException: Failed to create SaslClient
at org.apache.kafka.common.security.authenticator.SaslClientAuthenticator.createSaslClient(SaslClientAuthenticator.java:112)
at org.apache.kafka.common.security.authenticator.SaslClientAuthenticator.configure(SaslClientAuthenticator.java:94)
... 10 more
Caused by: javax.security.sasl.SaslException: PLAIN: authorization ID and password must be specified
at com.sun.security.sasl.PlainClient.<init>(PlainClient.java:58)
at com.sun.security.sasl.ClientFactoryImpl.createSaslClient(ClientFactoryImpl.java:97)
at javax.security.sasl.Sasl.createSaslClient(Sasl.java:384)
at com.ibm.messagehub.login.MessageHubSaslClientFactory.createSaslClient(MessageHubSaslClientFactory.java:77)
at javax.security.sasl.Sasl.createSaslClient(Sasl.java:384)
at org.apache.kafka.common.security.authenticator.SaslClientAuthenticator$1.run(SaslClientAuthenticator.java:107)
at org.apache.kafka.common.security.authenticator.SaslClientAuthenticator$1.run(SaslClientAuthenticator.java:102)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:422)
at org.apache.kafka.common.security.authenticator.SaslClientAuthenticator.createSaslClient(SaslClientAuthenticator.java:102)
... 11 more
```

4. The authentication failure causes the failure to create the KafkaChannel. The KafkaChannel obtained through the **channel(key)** method is empty and "NullPointerException" is excessively printed. According to the preceding log, the authentication fails due to an incorrect password which does not match the username.
5. Check the **jaas** and **keytab** files. The **principal** is set to **stream** in the **jaas** file.

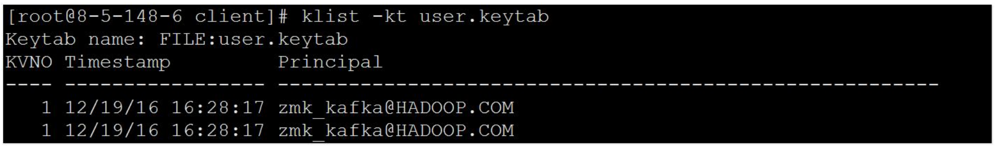
Figure 15-45 Checking the jaas file



```
KafkaClient {
com.sun.security.auth.module.Krb5LoginModule required
debug=false
keyTab="/opt/client/user.keytab"
useTicketCache=false
storeKey=true
principal="stream@HADOOP.COM"
useKeyTab=true;
};
```

The **principal** is set to **zmk_kafka** in the **user.keytab** file.

Figure 15-46 Checking the user.keytab file



```
[root@8-5-148-6 client]# klist -kt user.keytab
Keytab name: FILE:user.keytab
KVNO Timestamp         Principal
---- ----------------- --------------------------------------------------------
   1 12/19/16 16:28:17 zmk_kafka@HADOOP.COM
   1 12/19/16 16:28:17 zmk_kafka@HADOOP.COM
```

The **principal** in the **jaas** file is inconsistent with that in the **user.keytab** file.

The application automatically and periodically updates the **jaas** file. However, when two processes of the application update the **jaas** file, one process writes a correct **principal** whereas the other process writes an incorrect one. As a result, the application is abnormal sometimes.

Procedure

Step 1 Modify the **jaas** file to ensure that its **principal** exists in the **keytab** file.

----End

15.13.4 Producer Fails to Send Data and "TOPIC_AUTHORIZATION_FAILED" Is Thrown

Symptom

An MRS cluster is installed, and ZooKeeper and Kafka are installed in the cluster.

When Producer sends data to Kafka, the client throws "TOPIC_AUTHORIZATION_FAILED."

Possible Causes

1. The Kafka service is abnormal.
2. The Producer client adopts non-security access and access is disabled on the server.
3. The Producer client adopts non-security access and ACL is set for Kafka topics.

Cause Analysis

The possible reasons why Producer fails to send data to Kafka may be related to Producer or Kafka.

1. Check the Kafka service status:
   - MRS Manager: Log in to MRS Manager and choose **Services** > **Kafka**. Check the Kafka status. The status is **Good**, and the monitoring metrics are correctly displayed.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Kafka**.

     Check the Kafka status. It is found that the status is good and the monitoring metrics are correctly displayed.
2. Check the Producer client logs. The logs contain the error information "TOPIC_AUTHORIZATION_FAILED."
   ```
   [root@10-10-144-2 client]# kafka-console-producer.sh --broker-list 10.5.144.2:9092 --topic test
   1
   [2017-01-24 16:58:36,671] WARN Error while fetching metadata with correlation id 0 :
   {test=TOPIC_AUTHORIZATION_FAILED} (org.apache.kafka.clients.NetworkClient)
   [2017-01-24 16:58:36,672] ERROR Error when sending message to topic test with key: null, value: 1
   bytes with error: Not authorized to access topics: [test]
   (org.apache.kafka.clients.producer.internals.ErrorLoggingCallback)
   ```
   Producer accesses Kafka using port 9092, which is a non-security port.
3. On Manager, check the current Kafka cluster configuration. It is found that the customized configuration **allow.everyone.if.no.acl.found**=**false** is not configured.
   - MRS Manager portal: Log in to MRS Manager and choose **Services** > **Kafka** > **Configuration**.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Kafka** > **Configuration**.
4. If ACL is set to **false**, port 9092 cannot be used for access.
5. Check the Producer client logs. The logs contain the error information "TOPIC_AUTHORIZATION_FAILED."
   ```
   [root@10-10-144-2 client]# kafka-console-producer.sh --broker-list 10.5.144.2:21005 --topic test_acl
   1
   ```

```
[2017-01-25 11:09:40,012] WARN Error while fetching metadata with correlation id 0 :
{test_acl=TOPIC_AUTHORIZATION_FAILED} (org.apache.kafka.clients.NetworkClient)
[2017-01-25 11:09:40,013] ERROR Error when sending message to topic test_acl with key: null, value:
1 bytes with error: Not authorized to access topics: [test_acl]
(org.apache.kafka.clients.producer.internals.ErrorLoggingCallback)
[2017-01-25 11:14:40,010] WARN Error while fetching metadata with correlation id 1 :
{test_acl=TOPIC_AUTHORIZATION_FAILED} (org.apache.kafka.clients.NetworkClient)
```

Producer accesses Kafka using port 21005, which is a non-security port.

6. Run the client command to check the ACL permission of the topic.

   ```
   [root@10-10-144-2 client]# kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:24002/
   kafka --list --topic topic_acl
   Current ACLs for resource `Topic:topic_acl`:
    User:test_user has Allow permission for operations: Describe from hosts: *
   User:test_user has Allow permission for operations: Write from hosts: *
   ```

   If ACL is set for the topic, port 9092 cannot be used for access.

7. Check the Producer client logs. The logs contain the error information "TOPIC_AUTHORIZATION_FAILED."

   ```
   [root@10-10-144-2 client]# kafka-console-producer.sh --broker-list 10.5.144.2:21007 --topic topic_acl
   --producer.config /opt/client/Kafka/kafka/config/producer.properties
   1
   [2017-01-25 12:43:58,506] WARN Error while fetching metadata with correlation id 0 :
   {topic_acl=TOPIC_AUTHORIZATION_FAILED} (org.apache.kafka.clients.NetworkClient)
   [2017-01-25 12:43:58,507] ERROR Error when sending message to topic topic_acl with key: null,
   value: 1 bytes with error: Not authorized to access topics: [topic_acl]
   (org.apache.kafka.clients.producer.internals.ErrorLoggingCallback)
   ```

   Producer uses port 21007 to access Kafka.

8. Run the client command **klist** to query the current authenticated user.

   ```
   [root@10-10-144-2 client]# klist
   Ticket cache: FILE:/tmp/krb5cc_0
   Default principal: test@HADOOP.COM

   Valid starting     Expires            Service principal
   01/25/17 11:06:48  01/26/17 11:06:45  krbtgt/HADOOP.COM@HADOOP.COM
   ```

   The **test** user is used in this example.

9. Run the client command to check the ACL permission of the topic.

   ```
   [root@10-10-144-2 client]# kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:2181/
   kafka --list --topic topic_acl
   Current ACLs for resource `Topic:topic_acl`:
    User:test_user has Allow permission for operations: Describe from hosts: *
   User:test_user has Allow permission for operations: Write from hosts: *
   ```

   After ACL is set for the topic, user **test_user** has Producer permission. User **test** has no permission to perform Producer operations.

   For details about the solution, see 2.

10. Log in to Kafka Broker using SSH.

    Run the **cd /var/log/Bigdata/kafka/broker** command to go to the log directory.

    Check the **kafka-authorizer.log** file. It shows that the user does not belong to the **kafka** or **kafkaadmin** group.

    ```
    2017-01-25 13:26:33,648 | INFO  | [kafka-request-handler-0] | The principal is test, belongs to Group :
    [hadoop, ficommon] | kafka.authorizer.logger (SimpleAclAuthorizer.scala:169)
    2017-01-25 13:26:33,648 | WARN  | [kafka-request-handler-0] | The user is not belongs to kafka or
    kafkaadmin group, authorize failed! | kafka.authorizer.logger (SimpleAclAuthorizer.scala:170)
    ```

    For details about the solution, see 3.

Solution

Step 1 Set **allow.everyone.if.no.acl.found** to **true** and restart the Kafka service.

Step 2 Use the account with permission for login.

Example:

kinit test_user

Alternatively, grant the user with related permission.

NOTICE

This operation must be performed by the Kafka administrator (belonging to the **kafkaadmin** group).

Example:

kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:2181/kafka --topic topic_acl --producer --add --allow-principal User:test

```
[root@10-10-144-2 client]# kafka-acls.sh --authorizer-properties zookeeper.connect=8.5.144.2:2181/kafka --list --topic topic_acl
Current ACLs for resource `Topic:topic_acl`:
 User:test_user has Allow permission for operations: Describe from hosts: *
 User:test_user has Allow permission for operations: Write from hosts: *
  User:test has Allow permission for operations: Describe from hosts: *
  User:test has Allow permission for operations: Write from hosts: *
```

Step 3 Add the user to the **kafka** or **kafkaadmin** group.

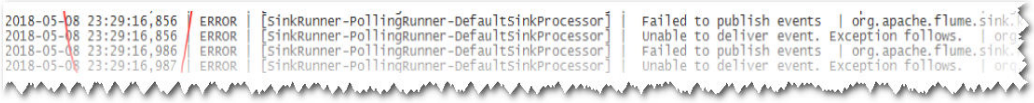
----End

15.13.5 Producer Occasionally Fails to Send Data and the Log Displays "Too many open files in system"

Symptom

When Producer sends data to Kafka, it is found that the client fails to send data.

Figure 15-47 Producer fails to send data.



```
2018-05-08 23:29:16,856 | ERROR | [SinkRunner-PollingRunner-DefaultSinkProcessor] | Failed to publish events  | org.apache.flume.sink.
2018-05-08 23:29:16,856 | ERROR | [SinkRunner-PollingRunner-DefaultSinkProcessor] | Unable to deliver event. Exception follows.  | org
2018-05-08 23:29:16,986 | ERROR | [SinkRunner-PollingRunner-DefaultSinkProcessor] | Failed to publish events  | org.apache.flume.sink.
2018-05-08 23:29:16,987 | ERROR | [SinkRunner-PollingRunner-DefaultSinkProcessor] | Unable to deliver event. Exception follows.  | org
```

Possible Causes

1. The Kafka service is abnormal.
2. The network is abnormal.
3. The Kafka topic is abnormal.

Cause Analysis

1. Check the Kafka service status:
   - MRS Manager: Log in to MRS Manager and choose **Services** > **Kafka**. Check the Kafka status. The status is **Good**, and the monitoring metrics are correctly displayed.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Kafka**.

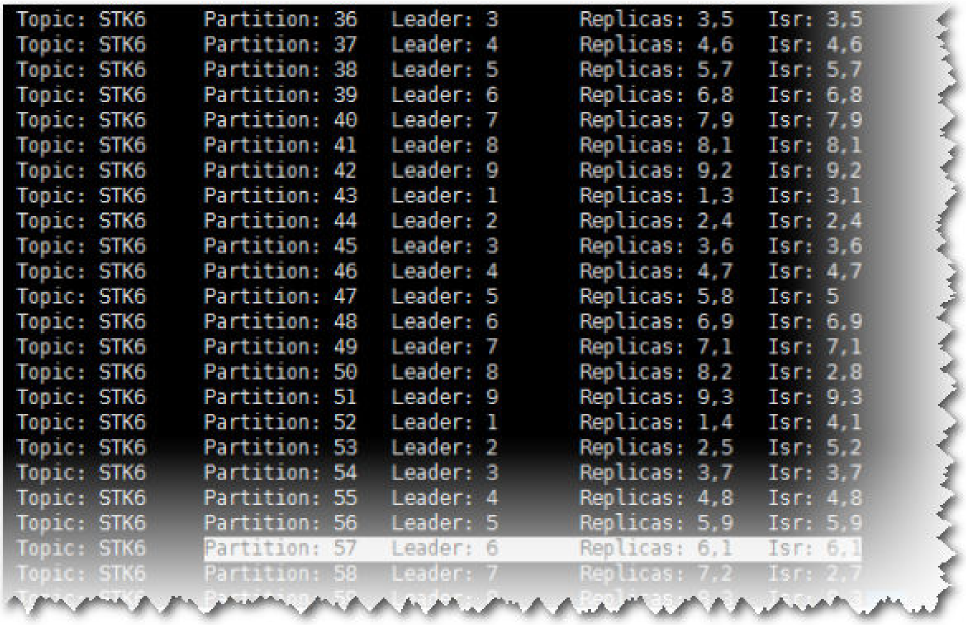
     Check the Kafka status. It is found that the status is good and the monitoring metrics are correctly displayed.
2. View the error topic information in the SparkStreaming log.

   Run the Kafka commands to obtain the topic assignment information and copy synchronization information, and check the return result.

   kafka-topics.sh --describe --zookeeper <zk_host:port/chroot>

   As shown in **Figure 15-48**, the topic status is normal. All partitions have normal leader information.

   Figure 15-48 Topic status

   ```
   Topic: STK6     Partition: 36  Leader: 3     Replicas: 3,5  Isr: 3,5
   Topic: STK6     Partition: 37  Leader: 4     Replicas: 4,6  Isr: 4,6
   Topic: STK6     Partition: 38  Leader: 5     Replicas: 5,7  Isr: 5,7
   Topic: STK6     Partition: 39  Leader: 6     Replicas: 6,8  Isr: 6,8
   Topic: STK6     Partition: 40  Leader: 7     Replicas: 7,9  Isr: 7,9
   Topic: STK6     Partition: 41  Leader: 8     Replicas: 8,1  Isr: 8,1
   Topic: STK6     Partition: 42  Leader: 9     Replicas: 9,2  Isr: 9,2
   Topic: STK6     Partition: 43  Leader: 1     Replicas: 1,3  Isr: 3,1
   Topic: STK6     Partition: 44  Leader: 2     Replicas: 2,4  Isr: 2,4
   Topic: STK6     Partition: 45  Leader: 3     Replicas: 3,6  Isr: 3,6
   Topic: STK6     Partition: 46  Leader: 4     Replicas: 4,7  Isr: 4,7
   Topic: STK6     Partition: 47  Leader: 5     Replicas: 5,8  Isr: 5
   Topic: STK6     Partition: 48  Leader: 6     Replicas: 6,9  Isr: 6,9
   Topic: STK6     Partition: 49  Leader: 7     Replicas: 7,1  Isr: 7,1
   Topic: STK6     Partition: 50  Leader: 8     Replicas: 8,2  Isr: 2,8
   Topic: STK6     Partition: 51  Leader: 9     Replicas: 9,3  Isr: 9,3
   Topic: STK6     Partition: 52  Leader: 1     Replicas: 1,4  Isr: 4,1
   Topic: STK6     Partition: 53  Leader: 2     Replicas: 2,5  Isr: 5,2
   Topic: STK6     Partition: 54  Leader: 3     Replicas: 3,7  Isr: 3,7
   Topic: STK6     Partition: 55  Leader: 4     Replicas: 4,8  Isr: 4,8
   Topic: STK6     Partition: 56  Leader: 5     Replicas: 5,9  Isr: 5,9
   Topic: STK6     Partition: 57  Leader: 6     Replicas: 6,1  Isr: 6,1
   Topic: STK6     Partition: 58  Leader: 7     Replicas: 7,2  Isr: 2,7
   ```
3. Run the **telnet** command to check whether the Kafka can be connected.

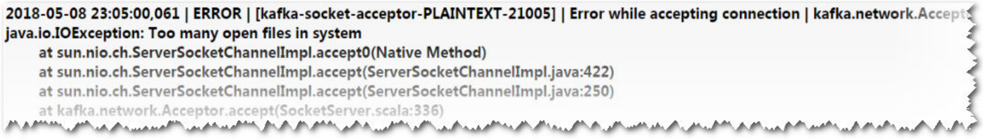
   telnet *Kafka service IP address Kafka service port*

   If telnet fails, check the network security group and ACL.
4. Log in to Kafka Broker using SSH.

   Run the **cd /var/log/Bigdata/kafka/broker** command to go to the log directory.

   Check on **server.log** indicates that the error message is displayed in the log shown in the following figure.

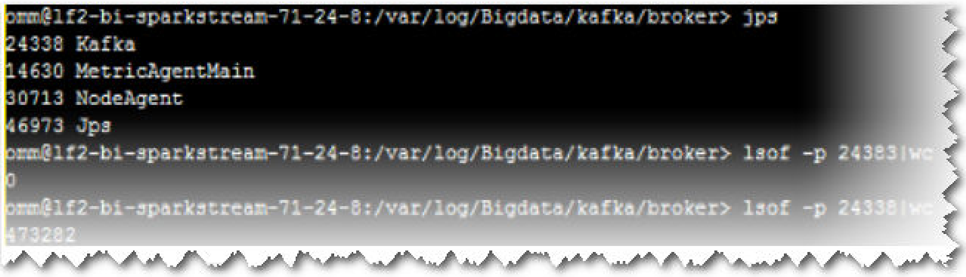
Figure 15-49 Log exception



```
2018-05-08 23:05:00,061 | ERROR | [kafka-socket-acceptor-PLAINTEXT-21005] | Error while accepting connection | kafka.network.Accept
java.io.IOException: Too many open files in system
        at sun.nio.ch.ServerSocketChannelImpl.accept0(Native Method)
        at sun.nio.ch.ServerSocketChannelImpl.accept(ServerSocketChannelImpl.java:422)
        at sun.nio.ch.ServerSocketChannelImpl.accept(ServerSocketChannelImpl.java:250)
        at kafka.network.Acceptor.accept(SocketServer.scala:336)
```

5. Output of the **lsof** command used to check the handle usage of the Kafka process on the current node shows that the number of handles used by the Kafka process reaches 470,000.

Figure 15-50 Handles



```
omm@1f2-bi-sparkstream-71-24-8:/var/log/Bigdata/kafka/broker> jps
24338 Kafka
14630 MetricAgentMain
30713 NodeAgent
46973 Jps
omm@1f2-bi-sparkstream-71-24-8:/var/log/Bigdata/kafka/broker> lsof -p 24383|wc
0
omm@1f2-bi-sparkstream-71-24-8:/var/log/Bigdata/kafka/broker> lsof -p 24338|wc
473282
```

6. Check the service code. It is found that the Producer object is frequently created and is not closed normally.

Solution

Step 1 Stop the current application to ensure that the number of handles on the server does not increase sharply, which affects the normal running of services.

Step 2 Optimize the application code to resolve the handle leakage problem.

Suggestion: Use one Producer object globally. After the use is complete, call the Close interface to close the handle.

----End

15.13.6 Consumer Is Initialized Successfully, But the Specified Topic Message Cannot Be Obtained from Kafka

Symptom

An MRS cluster is installed, and ZooKeeper, Flume, Kafka, Storm, and Spark are installed in the cluster.

The customer cannot consume any data using Storm, Spark, Flume or self-programmed Consumer code to consume messages of the specified Kafka topic.

Possible Causes

1. The Kafka service is abnormal.
2. The IP address for ZooKeeper connection is incorrectly set.
3. "ConsumerRebalanceFailedException" is thrown.

4. "ClosedChannelException" caused by network problems is thrown.

Cause Analysis

Storm, Spark, Flume or user-defined Consumer code can be called Consumer.

1. Check the Kafka service status:
   - MRS Manager: Log in to MRS Manager and choose **Services** > **Kafka**. Check the Kafka status. The status is **Good**, and the monitoring metrics are correctly displayed.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Kafka**.

     Check the Kafka status. It is found that the status is good and the monitoring metrics are correctly displayed.
2. Check whether data can be normally consumed through the Kafka client.

   Suppose the client has been installed in the **/opt/client** directory, **test** is the topic name to be consumed, and the IP address of ZooKeeper is 192.168.234.231.

   ```
   cd /opt/client
   source bigdata_env
   kinit admin
   kafka-topics.sh --zookeeper 192.168.234.231:2181/kafka --describe --topic testkafka-console-
   consumer.sh --topic test --zookeeper 192.168.234.231:2181/kafka --from-beginning
   ```

   If data can be consumed, the cluster service is running properly.
3. Check Consumer configurations. The IP address for connecting to ZooKeeper is incorrect.
   - Flume

     ```
     server.sources.Source02.type=org.apache.flume.source.kafka.KafkaSource

     server.sources.Source02.zookeeperConnect=192.168.234.231:2181
     server.sources.Source02.topic = test
     server.sources.Source02.groupId = test_01
     ```
   - Spark

     ```
     val zkQuorum = "192.168.234.231:2181"
     ```
   - Storm

     ```
     BrokerHosts brokerHosts = new ZKHosts("192.168.234.231:2181");
     ```
   - Consumer API

     ```
     zookeeper.connect="192.168.234.231:2181"
     ```

   On MRS Manager, the root path of ZNode where Kafka is stored on ZooKeeper is **/kafka**, which is differentiated from the open source. The address for Kafka to connect to ZooKeeper is **192.168.234.231:2181/kafka**.

   However, the address for Consumer to connect to ZooKeeper is **192.168.234.231:2181**. Therefore, topic information about Kafka cannot be correctly obtained.

   For details about the solution, see **Step 1**.
4. Check Consumer logs. The logs contain "ConsumerRebalanceFailedException".

   ```
   2016-02-03 15:55:32,557 | ERROR | [ZkClient-EventThread-75- 192.168.234.231:2181/kafka] |  Error
   handling event ZkEvent[New session event sent to kafka.consumer.ZookeeperConsumerConnector
   $ZKSessionExpireListener@34b41dfe]  | org.I0Itec.zkclient.ZkEventThread.run(ZkEventThread.java:77)
   kafka.common.ConsumerRebalanceFailedException: pc-zjqbetl86-1454482884879-2ec95ed3 can't
   rebalance after 4 retries
   at kafka.consumer.ZookeeperConsumerConnector
   $ZKRebalancerListener.syncedRebalance(ZookeeperConsumerConnector.scala:633)
   ```

```
at kafka.consumer.ZookeeperConsumerConnector
$ZKSessionExpireListener.handleNewSession(ZookeeperConsumerConnector.scala:487)
at org.I0Itec.zkclient.ZkClient$4.run(ZkClient.java:472)
at org.I0Itec.zkclient.ZkEventThread.run(ZkEventThread.java:71)
```

The exception shows that the current Consumer does not complete rebalance within the specified retry times. As a result, Kafka Topic-Partition is not allocated to Consumer and Consumer cannot consume messages.

For details about the solution, see **Step 3**.

5. Check Consumer logs. The logs contain "Fetching topic metadata with correlation id 0 for topics [Set(test)] from broker [id:26,host: 192-168-234-231,port:9092] failed" and "ClosedChannelException".

```
[2016-03-04 03:33:53,047] INFO Fetching metadata from broker id:26,host: 192-168-234-231,port:
9092 with correlation id 0 for 1 topic(s) Set(test) (kafka.client.ClientUtils$)
[2016-03-04 03:33:55,614] INFO Connected to 192-168-234-231:21005 for producing
(kafka.producer.SyncProducer)
[2016-03-04 03:33:55,614] INFO Disconnecting from 192-168-234-231:21005
(kafka.producer.SyncProducer)
[2016-03-04 03:33:55,615] WARN Fetching topic metadata with correlation id 0 for topics [Set(test)]
from broker [id:26,host: 192-168-234-231,port:21005] failed (kafka.client.ClientUtils$)
java.nio.channels.ClosedChannelException
at kafka.network.BlockingChannel.send(BlockingChannel.scala:100)
at kafka.producer.SyncProducer.liftedTree1$1(SyncProducer.scala:73)
at kafka.producer.SyncProducer.kafka$producer$SyncProducer$$doSend(SyncProducer.scala:72)
at kafka.producer.SyncProducer.send(SyncProducer.scala:113)
at kafka.client.ClientUtils$.fetchTopicMetadata(ClientUtils.scala:58)
at kafka.client.ClientUtils$.fetchTopicMetadata(ClientUtils.scala:93)
at kafka.consumer.ConsumerFetcherManager
$LeaderFinderThread.doWork(ConsumerFetcherManager.scala:66)
at kafka.utils.ShutdownableThread.run(ShutdownableThread.scala:60)
[2016-03-04 03:33:55,615] INFO Disconnecting from 192-168-234-231:21005
(kafka.producer.SyncProducer)
```

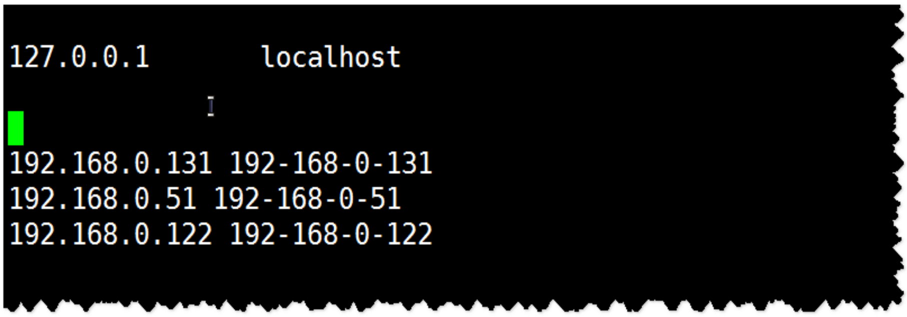
The exception shows that the current Consumer cannot obtain metadata from the Kafka Broker 192-168-234-231 node and cannot connect to the correct Broker for obtaining messages.

6. Check the network conditions. If the network is normal, check whether mapping between the host and the IP address is configured.

   – Linux

     Run the **cat /etc/hosts** command.

     Figure 15-51 Example 1

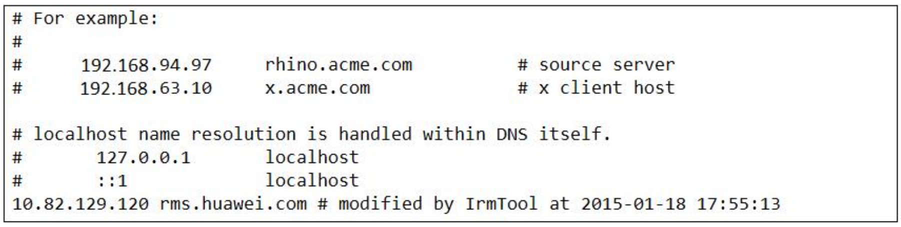


```
127.0.0.1       localhost

192.168.0.131 192-168-0-131
192.168.0.51 192-168-0-51
192.168.0.122 192-168-0-122
```

   – Windows

     Open **C:\Windows\System32\drivers\etc\hosts**.

Figure 15-52 Example 2



```
# For example:
#
#      192.168.94.97     rhino.acme.com          # source server
#      192.168.63.10     x.acme.com              # x client host

# localhost name resolution is handled within DNS itself.
#        127.0.0.1       localhost
#        ::1             localhost
10.82.129.120 rms.huawei.com # modified by IrmTool at 2015-01-18 17:55:13
```

For details about the solution, see **Step 4**.

Solution

Step 1 The IP address for connecting to ZooKeeper is incorrectly configured.

Step 2 Change the IP address for connecting to ZooKeeper in the Consumer configuration and make it consistent with MRS configuration.

- Flume

```
server.sources.Source02.type=org.apache.flume.source.kafka.KafkaSource
server.sources.Source02.zookeeperConnect=192.168.234.231:2181/kafka
server.sources.Source02.topic = test
server.sources.Source02.groupId = test_01
```

- Spark

```
val zkQuorum = "192.168.234.231:2181/kafka"
```

- Storm

```
BrokerHosts brokerHosts = new ZKHosts("192.168.234.231:2181/kafka");
```

- Consumer API

```
zookeeper.connect="192.168.234.231:2181/kafka"
```

Step 3 Rebalance is abnormal.

Multiple Consumers in the same consumer group are successively started and consume data of multiple partitions at the same time, load balancing is performed for Consumers when consumers are fewer than partitions.

The temporary node where the Consumer is stored on ZooKeeper determines read/write permission of which partition of which topic the Consumer has. The path is **/consumers/consumer-group-xxx/owners/topic-xxx/x**.

After the load balancing is triggered, the original Consumer will be recalculated and release occupied partitions, which takes a while. Therefore, new Consumers may fail to preempt the partitions.

Table 15-3 Parameters

Name	Function	Default Value
rebalance.max.retries	Maximum number of rebalance retries	4
rebalance.backoff.ms	Interval for each rebalance retry	2000

Name	Function	Default Value
zookeeper.session.timeout.ms	Maximum time allowed to create a session with ZooKeeper	15000

Set the preceding parameters to higher values. The following is for your reference:

```
zookeeper.session.timeout.ms = 45000
rebalance.max.retries = 10
rebalance.backoff.ms = 5000
```

Parameter setting must comply with the following rule:

rebalance.max.retries * rebalance.backoff.ms > zookeeper.session.timeout.ms

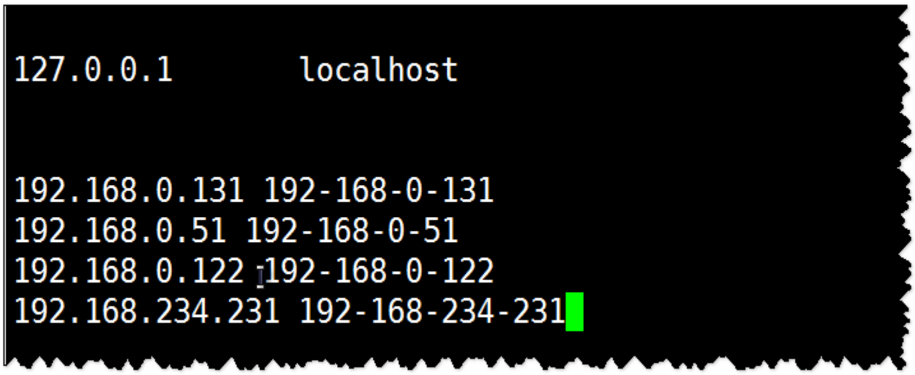
Step 4 The network is abnormal.

In the **hosts** file, mapping between the hostname and IP address is not configured. As a result, information cannot be obtained when using the hostname for access.

Step 5 Add the hostname to the **hosts** file and make it correspond to the IP address.

- Linux

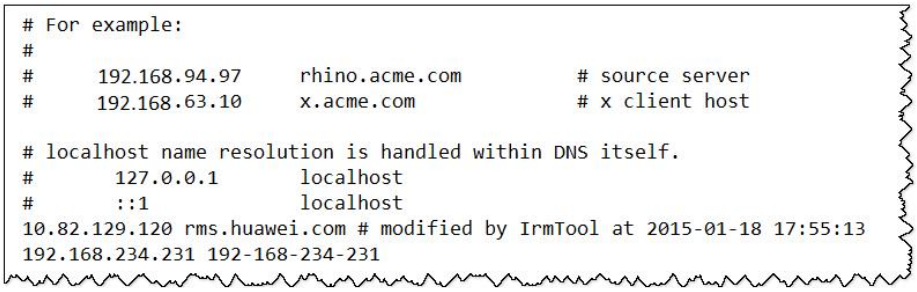
Figure 15-53 Example 3



```
127.0.0.1       localhost


192.168.0.131 192-168-0-131
192.168.0.51 192-168-0-51
192.168.0.122 192-168-0-122
192.168.234.231 192-168-234-231
```

- Windows

Figure 15-54 Example 4



```
# For example:
#
#      192.168.94.97     rhino.acme.com          # source server
#      192.168.63.10     x.acme.com              # x client host

# localhost name resolution is handled within DNS itself.
#       127.0.0.1       localhost
#       ::1             localhost
10.82.129.120 rms.huawei.com # modified by IrmTool at 2015-01-18 17:55:13
192.168.234.231 192-168-234-231
```

----End

15.13.7 Consumer Fails to Consume Data and Remains in the Waiting State

Symptom

An MRS cluster is installed, and ZooKeeper and Kafka are installed in the cluster.

When the Consumer consumes data from Kafka, the client stays in the Waiting state.

Possible Causes

1. The Kafka service is abnormal.
2. The Consumer client adopts non-security access and access is disabled on the server.
3. The Consumer client adopts non-security access and ACL is set for Kafka topics.

Cause Analysis

The possible reasons why the Consumer fails to consume data from Kafka may be related to the Consumer or Kafka.

1. Check the Kafka service status:
   - MRS Manager: Log in to MRS Manager and choose **Services** > **Kafka**. Check the Kafka status. The status is **Good**, and the monitoring metrics are correctly displayed.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Kafka**.

     Check the Kafka status. It is found that the status is good and the monitoring metrics are correctly displayed.
2. Check the Consumer client log. It is found that the information about the frequent connections and disconnections to the Broker node is printed, as shown in the following output.
   ```
   [root@10-10-144-2 client]# kafka-console-consumer.sh --topic test --zookeeper 10.5.144.2:2181/kafka
   --from-beginning


   [2017-03-07 09:22:00,658] INFO Fetching metadata from broker BrokerEndPoint(1,10.5.144.2,9092)
   with correlation id 26 for 1 topic(s) Set(test) (kafka.client.ClientUtils$)
   [2017-03-07 09:22:00,659] INFO Connected to 10.5.144.2:9092 for producing
   (kafka.producer.SyncProducer)
   [2017-03-07 09:22:00,659] INFO Disconnecting from 10.5.144.2:9092 (kafka.producer.SyncProducer)
   ```
   Consumer accesses Kafka using port 9092, which is a non-security port.
3. On Manager, check the current Kafka cluster configuration. It is found that the customized configuration **allow.everyone.if.no.acl.found**=**false** is not configured.
   - MRS Manager portal: Log in to MRS Manager and choose **Services** > **Kafka** > **Configuration**.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Kafka** > **Configuration**.

4. If ACL is set to **false**, port 9092 cannot be used for access.
5. Check the Consumer client log. It is found that the information about the frequent connections and disconnections to the Broker node is printed, as shown in the following output.
   ```
   [root@10-10-144-2 client]# kafka-console-consumer.sh --topic test_acl --zookeeper 10.5.144.2:2181/
   kafka --from-beginning


   [2017-03-07 09:49:16,992] INFO Fetching metadata from broker BrokerEndPoint(2,10.5.144.3,9092)
   with correlation id 16 for 1 topic(s) Set(topic_acl) (kafka.client.ClientUtils$)
   [2017-03-07 09:49:16,993] INFO Connected to 10.5.144.3:9092 for producing
   (kafka.producer.SyncProducer)
   [2017-03-07 09:49:16,994] INFO Disconnecting from 10.5.144.3:9092 (kafka.producer.SyncProducer)
   ```
   The Consumer accesses Kafka using port 21005, which is a non-security port.
6. Run the client command to check the ACL permission of the topic.
   ```
   [root@10-10-144-2 client]# kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:2181/
   kafka --list --topic topic_acl
   Current ACLs for resource `Topic:topic_acl`:
    User:test_user has Allow permission for operations: Describe from hosts: *
   User:test_user has Allow permission for operations: Write from hosts: *
   ```
   If ACL is set for the topic, port 9092 cannot be used for access.
7. The following information is printed in the Consumer client log:
   ```
   [root@10-10-144-2 client]# kafka-console-consumer.sh --topic topic_acl --bootstrap-server
   10.5.144.2:21007 --consumer.config /opt/client/Kafka/kafka/config/consumer.properties --from-
   beginning --new-consumer

   [2017-03-07 10:19:18,478] INFO Kafka version : 0.9.0.0 (org.apache.kafka.common.utils.AppInfoParser)
   [2017-03-07 10:19:18,478] INFO Kafka commitId : unknown
   (org.apache.kafka.common.utils.AppInfoParser)
   ```
   The Consumer uses port 21007 to access Kafka.
8. Run the client command **klist** to query the current authenticated user.
   ```
   [root@10-10-144-2 client]# klist
   Ticket cache: FILE:/tmp/krb5cc_0
   Default principal: test@HADOOP.COM

   Valid starting     Expires            Service principal
   01/25/17 11:06:48  01/26/17 11:06:45  krbtgt/HADOOP.COM@HADOOP.COM
   ```
   The **test** user is used in this example.
9. Run the client command to check the ACL permission of the topic.
   ```
   [root@10-10-144-2 client]# kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:24002/
   kafka --list --topic topic_acl
   Current ACLs for resource `Topic:topic_acl`:
    User:test_user has Allow permission for operations: Describe from hosts: *
    User:test_user has Allow permission for operations: Write from hosts: *
    User:ttest_user has Allow permission for operations: Read from hosts: *
   ```
   If ACL is set for the topic, user **test** does not have the permission to perform the Consumer operation.

   For details about the solution, see **Step 2**.
10. Log in to Kafka Broker using SSH.

    Run the **cd /var/log/Bigdata/kafka/broker** command to go to the log directory.

    Check the **kafka-authorizer.log** file. It shows that the user does not belong to the **kafka** or **kafkaadmin** group.
    ```
    2017-01-25 13:26:33,648 | INFO  | [kafka-request-handler-0] | The principal is test, belongs to Group :
    [hadoop, ficommon] | kafka.authorizer.logger (SimpleAclAuthorizer.scala:169)
    2017-01-25 13:26:33,648 | WARN  | [kafka-request-handler-0] | The user is not belongs to kafka or
    kafkaadmin group, authorize failed! | kafka.authorizer.logger (SimpleAclAuthorizer.scala:170)
    ```

For details about the solution, see **Step 3**.

Solution

Step 1 Set **allow.everyone.if.no.acl.found** to **true** and restart the Kafka service.

Step 2 Use the account with permission for login.

Example:

kinit test_user

Alternatively, grant the user with related permission.

NOTICE

This operation must be performed by the Kafka administrator (belonging to the **kafkaadmin** group).

Example:

kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:2181/kafka --topic topic_acl --consumer --add --allow-principal User:test --group test

```
[root@10-10-144-2 client]# kafka-acls.sh --authorizer-properties zookeeper.connect=8.5.144.2:2181/kafka --
list --topic topic_acl
Current ACLs for resource `Topic:topic_acl`:
 User:test_user has Allow permission for operations: Describe from hosts: *
 User:test_user has Allow permission for operations: Write from hosts: *
  User:test has Allow permission for operations: Describe from hosts: *
  User:test has Allow permission for operations: Write from hosts: *
  User:test has Allow permission for operations: Read from hosts: *
```

Step 3 Add the user to the **kafka** or **kafkaadmin** group.

----End

15.13.8 SparkStreaming Fails to Consume Kafka Messages, and "Error getting partition metadata" Is Displayed

Symptom

When SparkStreaming is used to consume messages of a specified topic in Kafka, data cannot be obtained from Kafka. The message "Error getting partition metadata" is displayed.

```
Exception in thread "main" org.apache.spark.SparkException:  Error getting partition metadata for
'testtopic'. Does the topic exist?
org.apache.spark.streaming.kafka.KafkaCluster$$anonfun$checkErrors$1.apply(KafkaCluster.scala:366)
org.apache.spark.streaming.kafka.KafkaCluster$$anonfun$checkErrors$1.apply(KafkaCluster.scala:366)
scala.util.Either.fold(Either.scala:97)
org.apache.spark.streaming.kafka.KafkaCluster$.checkErrors(KafkaCluster.scala:365)
org.apache.spark.streaming.kafka.KafkaUtils$.createDirectStream(KafkaUtils.scala:422)
com.XXXXXX.bigdata.spark.examples.FemaleInfoCollectionPrint$.main(FemaleInfoCollectionPrint.scala:45)
com.XXXXXX.bigdata.spark.examples.FemaleInfoCollectionPrint.main(FemaleInfoCollectionPrint.scala)
sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
```

```
java.lang.reflect.Method.invoke(Method.java:498)
org.apache.spark.deploy.SparkSubmit$.org$apache$spark$deploy$SparkSubmit$
$runMain(SparkSubmit.scala:762)
org.apache.spark.deploy.SparkSubmit$.doRunMain$1(SparkSubmit.scala:183)
org.apache.spark.deploy.SparkSubmit$.submit(SparkSubmit.scala:208)
org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:123)
org.apache.spark.deploy.SparkSubmit.main(SparkSubmit.scala)
```

Possible Causes

1. The Kafka service is abnormal.
2. The Consumer client adopts non-security access and access is disabled on the server.
3. The Consumer client adopts non-security access and ACL is set for Kafka topics.

Cause Analysis

1. Check the Kafka service status:
   - MRS Manager: Log in to MRS Manager and choose **Services** > **Kafka**. Check the Kafka status. The status is **Good**, and the monitoring metrics are correctly displayed.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Kafka**.

     Check the Kafka status. It is found that the status is good and the monitoring metrics are correctly displayed.
2. On Manager, check the current Kafka cluster configuration. It is found that **allow.everyone.if.no.acl.found** is not configured or is set to **false**.
   - MRS Manager portal: Log in to MRS Manager and choose **Services** > **Kafka** > **Configuration**.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Kafka** > **Configuration**.
3. If it is set to **false**, the Kafka non-secure port 21005 cannot be used for access.
4. Run the client command to check the ACL permission of the topic.

   ```
   [root@10-10-144-2 client]# kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:2181/
   kafka --list --topic topic_acl
   Current ACLs for resource `Topic:topic_acl`:
    User:test_user has Allow permission for operations: Describe from hosts: *
   User:test_user has Allow permission for operations: Write from hosts: *
   ```

   If an ACL is configured for a topic, the Kafka non-secure port 21005 cannot be used to access the topic.

Solution

Step 1 Add the customized configuration **allow.everyone.if.no.acl.found** or change its value to **true** and restart the Kafka service.

Step 2 Delete the ACL configured for the topic.

Example:

kinit test_user

NOTICE

This operation must be performed by the Kafka administrator (belonging to the **kafkaadmin** group).

Example:

kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:2181/kafka --remove --allow-principal User:test_user --producer --topic topic_acl

kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:2181/kafka --remove --allow-principal User:test_user --consumer --topic topic_acl --group test

----End

15.13.9 Consumer Fails to Consume Data in a Newly Created Cluster, and the Message " GROUP_COORDINATOR_NOT_AVAILABLE" Is Displayed

Symptom

A Kafka cluster is created, and two Broker nodes are deployed. The Kafka client can be used for production but cannot be used for consumption. The Consumer fails to consume data, and the message "GROUP_COORDINATOR_NOT_AVAILABLE" is displayed. The key log is as follows:

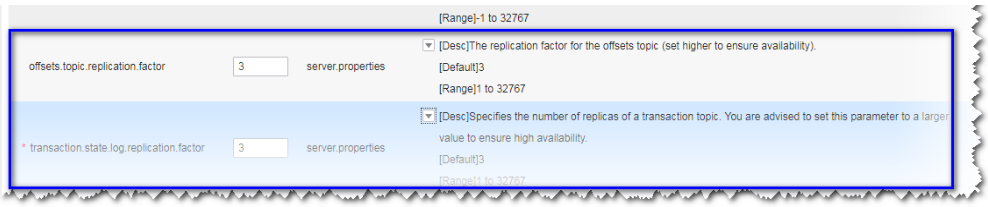
```
2018-05-12 10:58:42,561 | INFO | [kafka-request-handler-3] | [GroupCoordinator 2]: Preparing to restabilize group DemoConsumer with old generation 118 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
2018-05-12 10:59:13,562 | INFO | [executor-Heartbeat] | [GroupCoordinator 2]: Preparing to restabilize group DemoConsumer with old generation 119 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
```

Possible Causes

The **__consumer_offsets** cannot be created.

Cause Analysis

1. As indicated by the log, a large number of **_consumer_offset** creation operations failed.
2. The number of Brokers for the cluster is 2.
3. However, the number of replicas for the **_consumer_offset** topic is 3. Therefore, the creation fails.



			[Range]-1 to 32767
offsets.topic.replication.factor	3	server.properties	[Desc]The replication factor for the offsets topic (set higher to ensure availability). [Default]3 [Range]1 to 32767
* transaction.state.log.replication.factor	3	server.properties	[Desc]Specifies the number of replicas of a transaction topic. You are advised to set this parameter to a larger value to ensure high availability. [Default]3 [Range]1 to 32767

Solution

Expand the cluster to at least three streaming core nodes or perform the following steps to modify service configuration parameters:

Step 1 Go to the service configuration page.

- MRS Manager: Log in to MRS Manager, choose **Services** > **Kafka** > **Service Configuration**, and select **All** from **Type**.
- FusionInsight Manager: Log in to FusionInsight Manager. Choose **Cluster** > **Services** > **Kafka**. Click **Configurations** and select **All Configurations**.

Step 2 Search for **offsets.topic.replication.factor** and **transaction.state.log.replication.factor** and change their values to **2**.

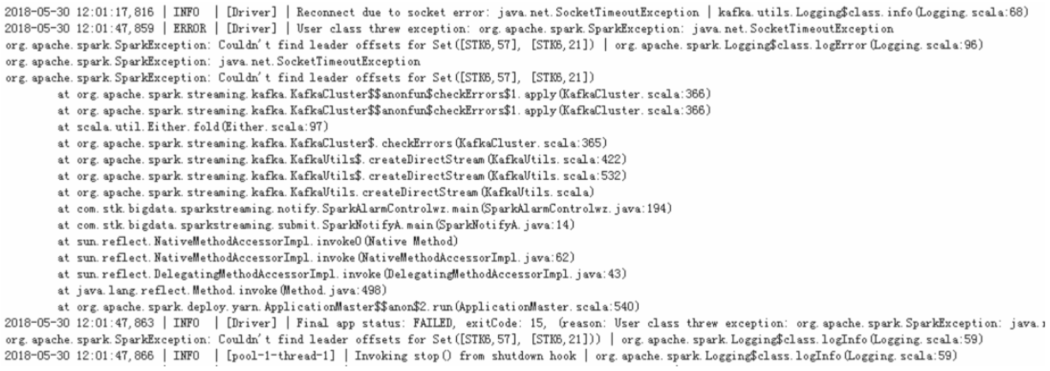
Step 3 Click **Save Configuration** and select **Restart the affected services or instances**. Click **OK** to restart the services.

----End

15.13.10 SparkStreaming Fails to Consume Kafka Messages, and the Message "Couldn't find leader offsets" Is Displayed

Symptom

When SparkStreaming is used to consume messages of a specified topic in Kafka, data cannot be obtained from Kafka. The following error message is displayed: Couldn't find leader offsets.



```
2018-05-30 12:01:17,816 | INFO  | [Driver] | Reconnect due to socket error: java.net.SocketTimeoutException | kafka.utils.Logging$class.info(Logging.scala:68)
2018-05-30 12:01:47,859 | ERROR | [Driver] | User class threw exception: org.apache.spark.SparkException: java.net.SocketTimeoutException
org.apache.spark.SparkException: Couldn't find leader offsets for Set([STK6,57], [STK6,21]) | org.apache.spark.Logging$class.logError(Logging.scala:96)
org.apache.spark.SparkException: java.net.SocketTimeoutException
org.apache.spark.SparkException: Couldn't find leader offsets for Set([STK6,57], [STK6,21])
        at org.apache.spark.streaming.kafka.KafkaCluster$$anonfun$checkErrors$1.apply(KafkaCluster.scala:366)
        at org.apache.spark.streaming.kafka.KafkaCluster$$anonfun$checkErrors$1.apply(KafkaCluster.scala:366)
        at scala.util.Either.fold(Either.scala:97)
        at org.apache.spark.streaming.kafka.KafkaCluster$.checkErrors(KafkaCluster.scala:365)
        at org.apache.spark.streaming.kafka.KafkaUtils$.createDirectStream(KafkaUtils.scala:422)
        at org.apache.spark.streaming.kafka.KafkaUtils$.createDirectStream(KafkaUtils.scala:532)
        at org.apache.spark.streaming.kafka.KafkaUtils.createDirectStream(KafkaUtils.scala)
        at com.stk.bigdata.sparkstreaming.notify.SparkAlarmControlwz.main(SparkAlarmControlwz.java:194)
        at com.stk.bigdata.sparkstreaming.submit.SparkNotifyA.main(SparkNotifyA.java:14)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.spark.deploy.yarn.ApplicationMaster$$anon$2.run(ApplicationMaster.scala:540)
2018-05-30 12:01:47,863 | INFO  | [Driver] | Final app status: FAILED, exitCode: 15, (reason: User class threw exception: org.apache.spark.SparkException: java.
org.apache.spark.SparkException: Couldn't find leader offsets for Set([STK6,57], [STK6,21])) | org.apache.spark.Logging$class.logInfo(Logging.scala:59)
2018-05-30 12:01:47,866 | INFO  | [pool-1-thread-1] | Invoking stop() from shutdown hook | org.apache.spark.Logging$class.logInfo(Logging.scala:59)
```

Possible Causes

- The Kafka service is abnormal.
- The network is abnormal.
- The Kafka topic is abnormal.

Cause Analysis

Step 1 On Manager, check the status of the Kafka cluster. The status is **Good**, and the monitoring metrics are correctly displayed.

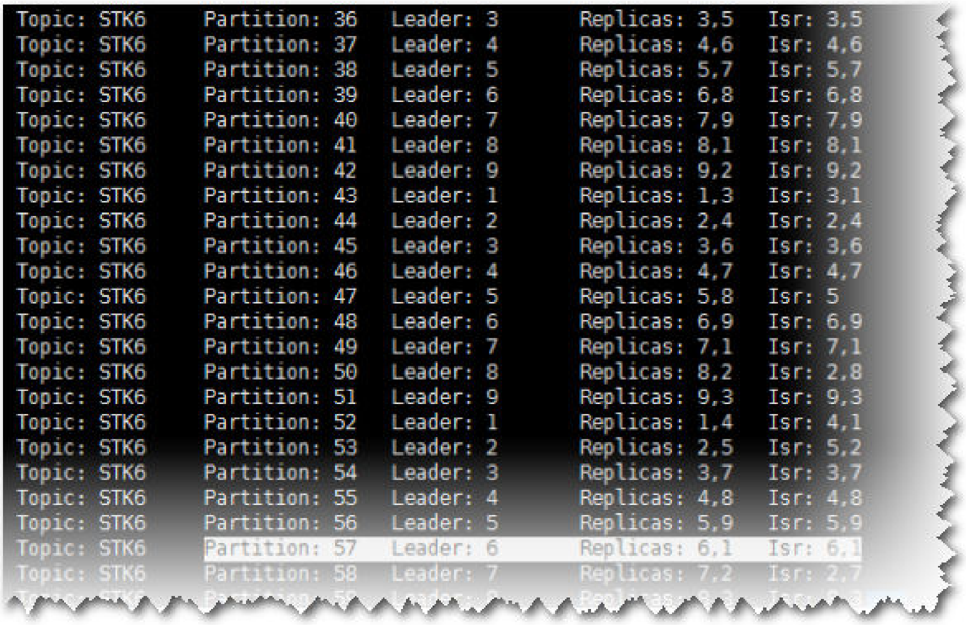
Step 2 View the error topic information in the SparkStreaming log.

Run the Kafka commands to obtain the topic assignment information and copy synchronization information, and check the return result.

kafka-topics.sh --describe --zookeeper <zk_host:port/chroot> --topic <topic name>

If information in the following figure is displayed, the topic is normal. All partitions have normal leader information.

Figure 15-55 Topic distribution information and copy synchronization information



```
Topic: STK6    Partition: 36   Leader: 3      Replicas: 3,5   Isr: 3,5
Topic: STK6    Partition: 37   Leader: 4      Replicas: 4,6   Isr: 4,6
Topic: STK6    Partition: 38   Leader: 5      Replicas: 5,7   Isr: 5,7
Topic: STK6    Partition: 39   Leader: 6      Replicas: 6,8   Isr: 6,8
Topic: STK6    Partition: 40   Leader: 7      Replicas: 7,9   Isr: 7,9
Topic: STK6    Partition: 41   Leader: 8      Replicas: 8,1   Isr: 8,1
Topic: STK6    Partition: 42   Leader: 9      Replicas: 9,2   Isr: 9,2
Topic: STK6    Partition: 43   Leader: 1      Replicas: 1,3   Isr: 3,1
Topic: STK6    Partition: 44   Leader: 2      Replicas: 2,4   Isr: 2,4
Topic: STK6    Partition: 45   Leader: 3      Replicas: 3,6   Isr: 3,6
Topic: STK6    Partition: 46   Leader: 4      Replicas: 4,7   Isr: 4,7
Topic: STK6    Partition: 47   Leader: 5      Replicas: 5,8   Isr: 5
Topic: STK6    Partition: 48   Leader: 6      Replicas: 6,9   Isr: 6,9
Topic: STK6    Partition: 49   Leader: 7      Replicas: 7,1   Isr: 7,1
Topic: STK6    Partition: 50   Leader: 8      Replicas: 8,2   Isr: 2,8
Topic: STK6    Partition: 51   Leader: 9      Replicas: 9,3   Isr: 9,3
Topic: STK6    Partition: 52   Leader: 1      Replicas: 1,4   Isr: 4,1
Topic: STK6    Partition: 53   Leader: 2      Replicas: 2,5   Isr: 5,2
Topic: STK6    Partition: 54   Leader: 3      Replicas: 3,7   Isr: 3,7
Topic: STK6    Partition: 55   Leader: 4      Replicas: 4,8   Isr: 4,8
Topic: STK6    Partition: 56   Leader: 5      Replicas: 5,9   Isr: 5,9
Topic: STK6    Partition: 57   Leader: 6      Replicas: 6,1   Isr: 6,1
Topic: STK6    Partition: 58   Leader: 7      Replicas: 7,2   Isr: 2,7
```

Step 3 Check whether the network connection between the client and Kafka cluster is normal. If no, contact the network team to rectify the fault.

Step 4 Log in to Kafka Broker using SSH.

Run the **cd /var/log/Bigdata/kafka/broker** command to go to the log directory.

Check on **server.log** indicates that the error message is displayed in the log shown in the following figure.

```
2018-05-30 12:02:00,246 | ERROR | [kafka-network-thread-6-PLAINTEXT-3] | Processor got uncaught
exception. | kafka.network.Processor (Logging.scala:103)

 java.lang.OutOfMemoryError: Direct buffer memory
 at java.nio.Bits.reserveMemory(Bits.java:694)
 at java.nio.DirectByteBuffer.<init>(DirectByteBuffer.java:123)
 at java.nio.ByteBuffer.allocateDirect(ByteBuffer.java:311)
 at sun.nio.ch.Util.getTemporaryDirectBuffer(Util.java:241)
 at sun.nio.ch.IOUtil.read(IOUtil.java:195)
 at sun.nio.ch.SocketChannelImpl.read(SocketChannelImpl.java:380)
```

at org.apache.kafka.common.network.PlaintextTransportLayer.read(PlaintextTransportLayer.java:110)

Step 5 On Manager, check the configuration of the current Kafka cluster.

- MRS Manager: Log in to MRS Manager and choose **Services** > **Kafka** > **Service Configuration**. Set **Type** to **All**. The value of **-XX:MaxDirectMemorySize** in **KAFKA_JVM_PERFORMANCE_OPTS** is **1G**.
- FusionInsight Manager: Log in to FusionInsight Manager. Choose **Cluster** > **Services** > **Kafka** > **Configurations** > **All Configurations**. The value of **-XX:MaxDirectMemorySize** in **KAFKA_JVM_PERFORMANCE_OPTS** is **1G**.

Step 6 If the direct memory is too small, an error is reported. Once the direct memory overflows, the node cannot process new requests. As a result, other nodes or clients fail to access the node due to timeout.

----End

Solution

Step 1 Log in to FusionInsight Manager and go to the Kafka configuration page.

- MRS Manager portal: Log in to MRS Manager and choose **Services** > **Kafka** > **Configuration**.
- FusionInsight Manager: Log in to FusionInsight Manager. Choose Cluster > **Services** > **Kafka** > **Configurations**.

Step 2 Set **Type** to **All**, and search for and change the value of **KAFKA_JVM_PERFORMANCE_OPTS**.

Step 3 Click **Save Configuration** and select **Restart the affected services or instances**. Click **OK** to restart the services.

----End

15.13.11 Consumer Fails to Consume Data and the Message " SchemaException: Error reading field 'brokers'" Is Displayed

Symptom

When a Consumer consumes messages of a specified topic in Kafka, the Consumer cannot obtain data from Kafka. The following error message is displayed: org.apache.kafka.common.protocol.types.SchemaException: Error reading field 'brokers': Error reading field 'host': Error reading string of length 28271, only 593 bytes available.

```
Exception in thread "Thread-0" org.apache.kafka.common.protocol.types.SchemaException: Error reading
field 'brokers': Error reading field 'host': Error reading string of length 28271, only 593 bytes available
at org.apache.kafka.common.protocol.types.Schema.read(Schema.java:73)
at org.apache.kafka.clients.NetworkClient.parseResponse(NetworkClient.java:380)
at org.apache.kafka.clients.NetworkClient.handleCompletedReceives(NetworkClient.java:449)
at org.apache.kafka.clients.NetworkClient.poll(NetworkClient.java:269)
at
org.apache.kafka.clients.consumer.internals.ConsumerNetworkClient.clientPoll(ConsumerNetworkClient.java:
360)
at org.apache.kafka.clients.consumer.internals.ConsumerNetworkClient.poll(ConsumerNetworkClient.java:
224)
at org.apache.kafka.clients.consumer.internals.ConsumerNetworkClient.poll(ConsumerNetworkClient.java:
192)
at org.apache.kafka.clients.consumer.internals.ConsumerNetworkClient.poll(ConsumerNetworkClient.java:
163)
atorg.apache.kafka.clients.consumer.internals.AbstractCoordinator.ensureCoordinatorReady(AbstractCoordina
tor.java:179)
at org.apache.kafka.clients.consumer.KafkaConsumer.pollOnce(KafkaConsumer.java:973)
at org.apache.kafka.clients.consumer.KafkaConsumer.poll(KafkaConsumer.java:937)
at KafkaNew.Consumer$ConsumerThread.run(Consumer.java:40)
```

Possible Causes

The JAR versions of the client and server are inconsistent.

Solution

Modify the Kafka JAR package in the Consumer application to ensure that it is the same as that on the server.

15.13.12 Checking Whether Data Consumed by a Customer Is Lost

Symptom

A Customer saves the consumed data to the database and finds that the data is inconsistent with the production data. Therefore, it is suspected that some of Kafka's consumed data is lost.

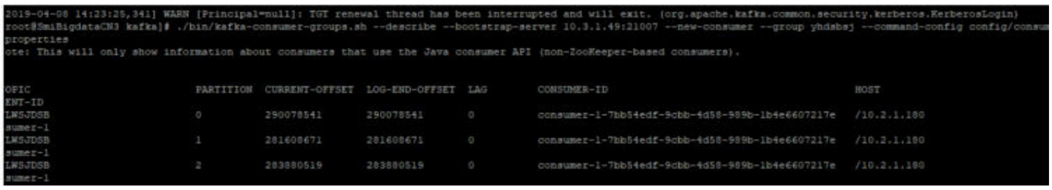
Possible Causes

- The customer code is incorrect.
- An exception occurs when Kafka production data is written.
- The Kafka consumption data is abnormal.

Solution

Check Kafka.

Step 1 Observe the changes of the written and consumed offset through **consumer-groups.sh**. (Produce a certain number of messages, and consume these messages on the client to observe the changes of the offset.)



```
2019-04-08 14:23:25,341] WARN [Principal=null]: TGT renewal thread has been interrupted and will exit. (org.apache.kafka.common.security.kerberos.KerberosLogin)
root@SmiBigdataCN3 kafka]# ./bin/kafka-consumer-groups.sh --describe --bootstrap-server 10.3.1.49:21007 --new-consumer --group yhdsbsj --command-config config/consum
properties
ote: This will only show information about consumers that use the Java consumer API (non-ZooKeeper-based consumers).


OPIC                    PARTITION  CURRENT-OFFSET  LOG-END-OFFSET  LAG        CONSUMER-ID                                       HOST
ENT-ID
LWSJDSB                 0          290078541       290078541       0          consumer-1-7bb54edf-9cbb-4d58-989b-1b4e6607217e   /10.2.1.180
sumer-1
LWSJDSB                 1          281608671       281608671       0          consumer-1-7bb54edf-9cbb-4d58-989b-1b4e6607217e   /10.2.1.180
sumer-1
LWSJDSB                 2          283880519       283880519       0          consumer-1-7bb54edf-9cbb-4d58-989b-1b4e6607217e   /10.2.1.180
sumer-1
```

Step 2 Create a consumption group, use the client to consume messages, and view the consumed messages.

new-consumer:

kafka-console-consumer.sh --topic <topic name> --bootstrap-server <IP1:PORT, IP2:PORT,...> --new-consumer --consumer.config <config file>

----End

Check the customer code.

Step 1 Check whether an error is reported when the offset is submitted on the client.

Step 2 If no error is reported, add a printing message to the API that is consumed, and print only the key to view the lost data.

----End

15.13.13 Failed to Start a Component Due to Account Lock

Symptom

In a new cluster, Kafka fails to be started. Authentication failure causes startup failure.

```
/home/omm/kerberos/bin/kinit -k -t /opt/XXXXXX/Bigdata/etc/2_15_ Broker /kafka.keytab kafka/
hadoop.hadoop.com -c /opt/XXXXXX/Bigdata/etc/2_15_ Broker /11846 failed.
export key tab file for kafka/hadoop.hadoop.com failed.export and check keytab file failed, errMsg=]}] for
Broker #192.168.1.92@192-168-1-92.
[2015-07-11 02:34:33] RoleInstance started failure for ROLE[name: Broker].
[2015-07-11 02:34:34] Failed to complete the instances start operation. Current operation entities: [Broker
#192.168.1.92@192-168-1-92], Failure entites : [Broker #192.168.1.92@192-168-1-92].Operation
Failed.Failed to complete the instances start operation. Current operation entities:
[Broker#192.168.1.92@192-168-1-92], Failure entites: [Broker #192.168.1.92@192-168-1-92].
```

Cause Analysis

Check the Kerberos log **/var/log/Bigdata/kerberos/krb5kdc.log**. It is found that IP addresses outside of the cluster uses the **kafka** account for connections, causing multiple authentication failures. As a result, the **kafka** account is locked.

```
Jul 11 02:49:16 192-168-1-91 krb5kdc[1863](info): AS_REQ (2 etypes {18 17}) 192.168.1.93:
NEEDED_PREAUTH: kafka/hadoop.hadoop.com@HADOOP.COM for krbtgt/HADOOP.COM@HADOOP.COM,
Additional pre-authentication required
Jul 11 02:49:16 192-168-1-91 krb5kdc[1863](info): preauth (encrypted_timestamp) verify failure: Decrypt
integrity check failed
Jul 11 02:49:16 192-168-1-91 krb5kdc[1863](info): AS_REQ (2 etypes {18 17}) 192.168.1.93:
PREAUTH_FAILED: kafka/hadoop.hadoop.com@HADOOP.COM for krbtgt/HADOOP.COM@HADOOP.COM,
Decrypt integrity check failed
```

Solution

Log in to a node outside the cluster (for example, 192.168.1.93 in the cause analysis example) and disable Kafka authentication. Wait 5 minutes for the account to be unlocked.

15.13.14 Kafka Broker Reports Abnormal Processes and the Log Shows "IllegalArgumentException"

Symptom

The Process Fault alarm is reported on Manager. Check whether the faulty process is Kafka Broker.

Possible Causes

Broker configuration is abnormal.

Cause Analysis

1. On Manager, obtain the host information on the alarm page.
2. Log in to Kafka Broker using SSH. Run the **cd /var/log/Bigdata/kafka/broker** command to go to the log directory.

   Check the **server.log** file. It is found that the "IllegalArgumentException" exception is thrown in the following log stating "java.lang.IllegalArgumentException: requirement failed: replica.fetch.max.bytes should be equal or greater than message.max.bytes."

   ```
   2017-01-25 09:09:14,930 | FATAL | [main] |  | kafka.Kafka$ (Logging.scala:113)
   java.lang.IllegalArgumentException: requirement failed: replica.fetch.max.bytes should be equal or
   greater than message.max.bytes
           at scala.Predef$.require(Predef.scala:233)
           at kafka.server.KafkaConfig.validateValues(KafkaConfig.scala:959)
           at kafka.server.KafkaConfig.<init>(KafkaConfig.scala:944)
           at kafka.server.KafkaConfig$.fromProps(KafkaConfig.scala:701)
           at kafka.server.KafkaConfig$.fromProps(KafkaConfig.scala:698)
           at kafka.server.KafkaServerStartable$.fromProps(KafkaServerStartable.scala:28)
           at kafka.Kafka$.main(Kafka.scala:60)
           at kafka.Kafka.main(Kafka.scala)
   ```

   Kafka requires that **replica.fetch.max.bytes** be greater than or equal to **message.max.bytes**.
3. On the Kafka configuration page, select **All Configurations**. All Kafka configurations are displayed. Search for **message.max.bytes** and

   replica.fetch.max.bytes. It is found that the value of **replica.fetch.max.bytes** is less than that of **message.max.bytes**.

Solution

Step 1 Go to the Kafka configuration page.

- For versions earlier than MRS 3.*x*: Log in to MRS Manager and choose **Services** > **Kafka** > **Service Configuration** > **All Configurations**.
- For MRS 3.*x* or later: Log in to FusionInsight Manager and choose **Cluster** > **Services** > **Kafka** > **Configurations** > **All Configurations**.

Step 2 Search for and modify the **replica.fetch.max.bytes** parameter to ensure that its value is greater than or equal to that of **message.max.bytes**. In this way, replicas of partitions on different brokers can be synchronized to all messages.

Step 3 Save the configuration and check whether there is any service whose configuration has expired in the cluster. If yes, restart the corresponding service or role instance for the configuration to take effect.

Step 4 Modify **fetch.message.max.bytes** in the Consumer service application to ensure that the value of **fetch.message.max.bytes** is greater than or equal to that of **message.max.bytes**.

----End

15.13.15 Kafka Topics Cannot Be Deleted

Symptom

When running the following command on the Kafka client to delete topics, it is found that the topics cannot be deleted.

```
kafka-topics.sh --delete --topic test --zookeeper 192.168.234.231:2181/kafka
```

Possible Causes

- The command for connecting the client to ZooKeeper is incorrect.
- Kafka is abnormal and some Kafka nodes are stopped.
- Perform the following operations when Kafka server configurations cannot be deleted.
- Perform the following operations when Kafka configurations are automatically created and the Producer is not stopped.

Cause Analysis

1. After the client command is run, the "ZkTimeoutException" exception is reported.

```
[2016-03-09 10:41:45,773] WARN Can not get the principle name from server 192.168.234.231
(org.apache.zookeeper.ClientCnxn)
Exception in thread "main" org.I0Itec.zkclient.exception.ZkTimeoutException: Unable to connect to
zookeeper server within timeout: 30000
at org.I0Itec.zkclient.ZkClient.connect(ZkClient.java:880)
at org.I0Itec.zkclient.ZkClient.<init>(ZkClient.java:98)
at org.I0Itec.zkclient.ZkClient.<init>(ZkClient.java:84)
at kafka.admin.TopicCommand$.main(TopicCommand.scala:51)
at kafka.admin.TopicCommand.main(TopicCommand.scala)
```

   For details about the solution, see **Step 1**.

2. Run the following query command on the client:

   kafka-topics.sh --list --zookeeper 192.168.0.122:2181/kafka

```
test - marked for deletion
```

   On Manager, check the running status of Kafka Broker instances.

   Run the **cd /var/log/Bigdata/kafka/broker** command to go to the log directory of node **RunningAsController**. Locate **ineligible for deletion: test** in the **controller.log** file.

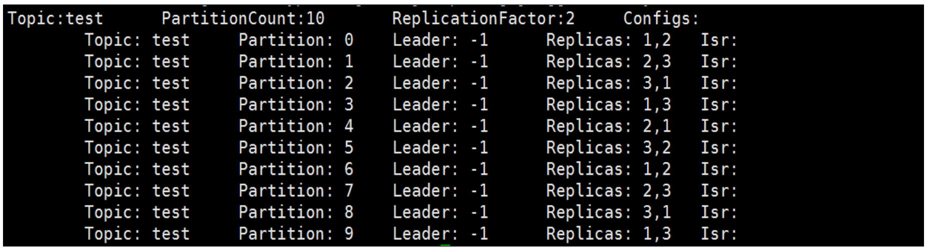
```
2016-03-09 11:11:26,228 | INFO  | [main] | [Controller 1]: List of topics to be deleted:  |
kafka.controller.KafkaController (Logging.scala:68)
2016-03-09 11:11:26,229 | INFO  | [main] | [Controller 1]: List of topics ineligible for deletion: test |
kafka.controller.KafkaController (Logging.scala:68)
```

3. On Manager, view the **delete.topic.enable** status of Broker.

   For details about the solution, see **Step 2**.

4. Run the following query command on the client:

```
kafka-topics.sh --describe –topic test --zookeeper 192.168.0.122:2181/kafka
```



```
Topic:test      PartitionCount:10       ReplicationFactor:2     Configs:
        Topic: test     Partition: 0    Leader: -1      Replicas: 1,2   Isr:
        Topic: test     Partition: 1    Leader: -1      Replicas: 2,3   Isr:
        Topic: test     Partition: 2    Leader: -1      Replicas: 3,1   Isr:
        Topic: test     Partition: 3    Leader: -1      Replicas: 1,3   Isr:
        Topic: test     Partition: 4    Leader: -1      Replicas: 2,1   Isr:
        Topic: test     Partition: 5    Leader: -1      Replicas: 3,2   Isr:
        Topic: test     Partition: 6    Leader: -1      Replicas: 1,2   Isr:
        Topic: test     Partition: 7    Leader: -1      Replicas: 2,3   Isr:
        Topic: test     Partition: 8    Leader: -1      Replicas: 3,1   Isr:
        Topic: test     Partition: 9    Leader: -1      Replicas: 1,3   Isr:
```

   Go to the log directory of node **RunningAsController**. Locate **marked ineligible for deletion** in the **controller.log** file.

```
2016-03-10 11:11:17,989 | INFO  | [delete-topics-thread-3] | [delete-topics-thread-3], Handling
deletion for topics test | kafka.controller.TopicDeletionManager$DeleteTopicsThread (Logging.scala:68)
2016-03-10 11:11:17,990 | INFO  | [delete-topics-thread-3] | [delete-topics-thread-3], Not retrying
```

```
deletion of topic test at this time since it is marked ineligible for deletion |
kafka.controller.TopicDeletionManager$DeleteTopicsThread (Logging.scala:68)
```

5. On Manager, query the Broker status.

   It can be seen that a Broker is in the Stopped state. In this case, delete the topic and ensure that Brokers where partitions of the topic reside must be in the Good state.

   For details about the solution, see **Step 3**.

6. Go to the log directory of node **RunningAsController**. Locate **Deletion successfully** in the **controller.log** file. If **New topics:[Set(test)]** is displayed again, it indicates that the topic is created again.

```
2016-03-10 11:33:35,208 | INFO  | [delete-topics-thread-3] | [delete-topics-thread-3], Deletion of topic
test successfully completed | kafka.controller.TopicDeletionManager$DeleteTopicsThread
(Logging.scala:68)

2016-03-10 11:33:38,501 | INFO  | [ZkClient-
EventThread-19-192.168.0.122:2181,160.172.0.52:2181,160.172.0.51:2181/kafka] |
[TopicChangeListener on Controller 3]: New topics: [Set(test)], deleted topics: [Set()], new partition
replica assignment
```

7. Use Manager to query the topic creation configuration of Broker.

   It is confirmed that the application that performs operations on the topic is not stopped.

   For details about the solution, see **Step 4**.

Solution

Step 1 Perform the following operations when connection to ZooKeeper fails.

When the connection between the Kafka client and ZooKeeper times out, run the ping command to check whether the Kafka client can connect to ZooKeeper. Check the network connection between the client and ZooKeeper.

If the network connection fails, check the ZooKeeper service information on Manager.

If ZooKeeper is improperly configured, change the ZooKeeper IP address in the client command.

Step 2 Perform the following operations when Kafka server configurations cannot be deleted.

On Manager, change the value of **delete.topic.enable** to **true**. Save the configurations and restart the service.

The client query command does not contain **Topic:test**.

```
kafka-topics.sh --list --zookeeper 192.168.0.122:24002/kafka
```

Go to the log directory of node **RunningAsController**. Locate **Deletion of topic test successfully** in the **controller.log** file.

```
2016-03-10 10:39:40,665 | INFO  | [delete-topics-thread-3] | [Partition state machine on Controller 3]:
Invoking state change to OfflinePartition for partitions [test,2],[test,15],[test,6],[test,16],[test,12],[test,7],
[test,10],[test,13],[test,9],[test,19],[test,3],[test,5],[test,1],[test,0],[test,17],[test,8],[test,4],[test,11],[test,14],
[test,18] | kafka.controller.PartitionStateMachine (Logging.scala:68)
2016-03-10 10:39:40,668 | INFO  | [delete-topics-thread-3] | [Partition state machine on Controller 3]:
Invoking state change to NonExistentPartition for partitions [test,2],[test,15],[test,6],[test,16],[test,12],[test,
7],[test,10],[test,13],[test,9],[test,19],[test,3],[test,5],[test,1],[test,0],[test,17],[test,8],[test,4],[test,11],[test,
14],[test,18] | kafka.controller.PartitionStateMachine (Logging.scala:68)
```

```
2016-03-10 10:39:40,977 | INFO  | [delete-topics-thread-3] | [delete-topics-thread-3], Deletion of topic test
successfully completed | kafka.controller.TopicDeletionManager$DeleteTopicsThread (Logging.scala:68)
```

Step 3 Some Kafka nodes are stopped or faulty.

Start the stopped Broker instances.

The client query command does not contain **Topic:test**.

```
kafka-topics.sh --list --zookeeper 192.168.0.122:24002/kafka
```

Go to the log directory of node **RunningAsController**. Locate **Deletion of topic test successfully** in the **controller.log** file.

```
2016-03-10 11:17:56,463 | INFO  | [delete-topics-thread-3] | [Partition state machine on Controller 3]:
Invoking state change to NonExistentPartition for partitions [test,4],[test,1],[test,8],[test,2],[test,5],[test,9],
[test,7],[test,6],[test,0],[test,3] | kafka.controller.PartitionStateMachine (Logging.scala:68)
2016-03-10 11:17:56,726 | INFO  | [delete-topics-thread-3] | [delete-topics-thread-3], Deletion of topic test
successfully completed | kafka.controller.TopicDeletionManager$DeleteTopicsThread (Logging.scala:68)
```

Step 4 Perform the following operations when Kafka configurations are automatically created and the Producer is not stopped.

Stop related applications, change the value of **auto.create.topics.enable** to **false** on Manager. Save the configuration and restart the service.

Step 5 Perform the delete operation again.

----End

15.13.16 Error "AdminOperationException" Is Displayed When a Kafka Topic Is Deleted

Symptom

When running the following command on the Kafka client to set the ACL for a topic, it is found that the ACL cannot be set.

```
kafka-topics.sh --delete --topic test4 --zookeeper 10.5.144.2:2181/kafka
```

The error message "ERROR kafka.admin.AdminOperationException: Error while deleting topic test4" is displayed.

Details are as follows:

```
Error while executing topic command : Error while deleting topic test4
[2017-01-25 14:00:20,750] ERROR kafka.admin.AdminOperationException: Error while deleting topic test4
at kafka.admin.TopicCommand$$anonfun$deleteTopic$1.apply(TopicCommand.scala:177)
at kafka.admin.TopicCommand$$anonfun$deleteTopic$1.apply(TopicCommand.scala:162)
at scala.collection.mutable.ResizableArray$class.foreach(ResizableArray.scala:59)
at scala.collection.mutable.ArrayBuffer.foreach(ArrayBuffer.scala:47)
at kafka.admin.TopicCommand$.deleteTopic(TopicCommand.scala:162)
at kafka.admin.TopicCommand$.main(TopicCommand.scala:68)
at kafka.admin.TopicCommand.main(TopicCommand.scala)
 (kafka.admin.TopicCommand$)
```

Possible Causes

The user does not belong to the **kafkaadmin** group. Kafka provides a secure access interface. Only users in the **kafkaadmin** group can delete topics.

Cause Analysis

1. After the client command is run, the "AdminOperationException" exception is reported.
2. Run the client command **klist** to query the current authenticated user.

   ```
   [root@10-10-144-2 client]# klist
   Ticket cache: FILE:/tmp/krb5cc_0
   Default principal: test@HADOOP.COM

   Valid starting     Expires            Service principal
   01/25/17 11:06:48  01/26/17 11:06:45  krbtgt/HADOOP.COM@HADOOP.COM
   ```

   The **test** user is used in this example.
3. Run the **id** command to query the user group information.

   ```
   [root@10-10-144-2 client]# id test
   uid=20032(test) gid=10001(hadoop) groups=10001(hadoop),9998(ficommon),10003(kafka)
   ```

Solution

MRS Manager:

Step 1 Log in to MRS Manager.

Step 2 Choose **System** > **Manage User**.

Step 3 In the **Operation** column of the user, click **Modify**.

Step 4 Add the user to the **kafkaadmin** group. Click **OK**.

Step 5 Run the **id** command to query the user group information.

```
[root@10-10-144-2 client]# id test
uid=20032(test) gid=10001(hadoop) groups=10001(hadoop),9998(ficommon),10002(kafkaadmin),
10003(kafka)
```

----End

FusionInsight Manager:

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 Locate the row that contains the target user, and click **Modify**.

Step 4 Add the user to the **kafkaadmin** group. Click **OK**.

Step 5 Run the **id** command to query the user group information.

```
[root@10-10-144-2 client]# id test
uid=20032(test) gid=10001(hadoop) groups=10001(hadoop),9998(ficommon),10002(kafkaadmin),
10003(kafka)
```

----End

15.13.17 When a Kafka Topic Fails to Be Created, "NoAuthException" Is Displayed

Symptom

When running the following command on the Kafka client to create topics, it is found that the topics cannot be created.

```
kafka-topics.sh --create --zookeeper 192.168.234.231:2181/kafka --replication-factor 1 --partitions 2 --
topic test
```

Error messages "NoAuthException" and "KeeperErrorCode = NoAuth for /config/topics" are displayed.

Details are as follows:

```
Error while executing topic command org.apache.zookeeper.KeeperException$NoAuthException:
KeeperErrorCode = NoAuth for /config/topics
org.I0Itec.zkclient.exception.ZkException: org.apache.zookeeper.KeeperException$NoAuthException:
KeeperErrorCode = NoAuth for /config/topics
 at org.I0Itec.zkclient.exception.ZkException.create(ZkException.java:68)
 at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:685)
 at org.I0Itec.zkclient.ZkClient.create(ZkClient.java:304)
 at org.I0Itec.zkclient.ZkClient.createPersistent(ZkClient.java:213)
 at kafka.utils.ZkUtils$.createParentPath(ZkUtils.scala:215)
 at kafka.utils.ZkUtils$.updatePersistentPath(ZkUtils.scala:338)
 at kafka.admin.AdminUtils$.writeTopicConfig(AdminUtils.scala:247)
```

Possible Causes

The user does not belong to the **kafkaadmin** group. Kafka provides a secure access interface. Only users in the **kafkaadmin** group can delete topics.

Cause Analysis

1. After the client command is run, the "NoAuthException" exception is reported.

   ```
   Error while executing topic command org.apache.zookeeper.KeeperException$NoAuthException:
   KeeperErrorCode = NoAuth for /config/topics
   org.I0Itec.zkclient.exception.ZkException: org.apache.zookeeper.KeeperException$NoAuthException:
   KeeperErrorCode = NoAuth for /config/topics
    at org.I0Itec.zkclient.exception.ZkException.create(ZkException.java:68)
    at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:685)
    at org.I0Itec.zkclient.ZkClient.create(ZkClient.java:304)
    at org.I0Itec.zkclient.ZkClient.createPersistent(ZkClient.java:213)
    at kafka.utils.ZkUtils$.createParentPath(ZkUtils.scala:215)
    at kafka.utils.ZkUtils$.updatePersistentPath(ZkUtils.scala:338)
    at kafka.admin.AdminUtils$.writeTopicConfig(AdminUtils.scala:247)
   ```

2. Run the client command **klist** to query the current authenticated user.

   ```
   [root@10-10-144-2 client]# klist
   Ticket cache: FILE:/tmp/krb5cc_0
   Default principal: test@HADOOP.COM

   Valid starting     Expires            Service principal
   01/25/17 11:06:48  01/26/17 11:06:45  krbtgt/HADOOP.COM@HADOOP.COM
   ```

   The **test** user is used in this example.

3. Run the **id** command to query the user group information.

   ```
   [root@10-10-144-2 client]# id test
   uid=20032(test) gid=10001(hadoop) groups=10001(hadoop),9998(ficommon),10003(kafka)
   ```

Solution

MRS Manager:

Step 1 Log in to MRS Manager.

Step 2 Choose **System** > **Manage User**.

Step 3 In the **Operation** column of the user, click **Modify**.

Step 4 Add the user to the **kafkaadmin** group.

Step 5 Run the **id** command to query the user group information.

```
[root@10-10-144-2 client]# id test
uid=20032(test) gid=10001(hadoop) groups=10001(hadoop),9998(ficommon),10002(kafkaadmin),
10003(kafka)
```

----End

FusionInsight Manager:

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 Locate the row that contains the target user, and click **Modify**.

Step 4 Add the user to the **kafkaadmin** group. Click **OK**.

Step 5 Run the **id** command to query the user group information.

```
[root@10-10-144-2 client]# id test
uid=20032(test) gid=10001(hadoop) groups=10001(hadoop),9998(ficommon),10002(kafkaadmin),
10003(kafka)
```

----End

15.13.18 Failed to Set an ACL for a Kafka Topic, and "NoAuthException" Is Displayed

Symptom

When running the following command on the Kafka client to set the ACL for a topic, it is found that the topic ACL cannot be set.

```
kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:2181/kafka  --topic topic_acl --producer
--add --allow-principal User:test_acl
```

The error message "NoAuthException: KeeperErrorCode = NoAuth for /kafka-acl-changes/acl_changes_0000000002" is displayed.

Details are as follows:

```
Error while executing ACL command: org.apache.zookeeper.KeeperException$NoAuthException:
KeeperErrorCode = NoAuth for /kafka-acl-changes/acl_changes_0000000002
org.I0Itec.zkclient.exception.ZkException: org.apache.zookeeper.KeeperException$NoAuthException:
KeeperErrorCode = NoAuth for /kafka-acl-changes/acl_changes_0000000002
at org.I0Itec.zkclient.exception.ZkException.create(ZkException.java:68)
at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:995)
at org.I0Itec.zkclient.ZkClient.delete(ZkClient.java:1038)
at kafka.utils.ZkUtils.deletePath(ZkUtils.scala:499)
at kafka.common.ZkNodeChangeNotificationListener$$anonfun$purgeObsoleteNotifications
$1.apply(ZkNodeChangeNotificationListener.scala:118)
at kafka.common.ZkNodeChangeNotificationListener$$anonfun$purgeObsoleteNotifications
$1.apply(ZkNodeChangeNotificationListener.scala:112)
at scala.collection.mutable.ResizableArray$class.foreach(ResizableArray.scala:59)
at scala.collection.mutable.ArrayBuffer.foreach(ArrayBuffer.scala:47)
at
kafka.common.ZkNodeChangeNotificationListener.purgeObsoleteNotifications(ZkNodeChangeNotificationLis
tener.scala:112)
at kafka.common.ZkNodeChangeNotificationListener.kafka$common$ZkNodeChangeNotificationListener$
$processNotifications(ZkNodeChangeNotificationListener.scala:97)
at
kafka.common.ZkNodeChangeNotificationListener.processAllNotifications(ZkNodeChangeNotificationListene
r.scala:77)
at kafka.common.ZkNodeChangeNotificationListener.init(ZkNodeChangeNotificationListener.scala:65)
at kafka.security.auth.SimpleAclAuthorizer.configure(SimpleAclAuthorizer.scala:136)
at kafka.admin.AclCommand$.withAuthorizer(AclCommand.scala:73)
at kafka.admin.AclCommand$.addAcl(AclCommand.scala:80)
at kafka.admin.AclCommand$.main(AclCommand.scala:48)
at kafka.admin.AclCommand.main(AclCommand.scala)
Caused by: org.apache.zookeeper.KeeperException$NoAuthException: KeeperErrorCode = NoAuth for /kafka-
acl-changes/acl_changes_0000000002
at org.apache.zookeeper.KeeperException.create(KeeperException.java:117)
at org.apache.zookeeper.KeeperException.create(KeeperException.java:51)
at org.apache.zookeeper.ZooKeeper.delete(ZooKeeper.java:1416)
at org.I0Itec.zkclient.ZkConnection.delete(ZkConnection.java:104)
at org.I0Itec.zkclient.ZkClient$11.call(ZkClient.java:1042)
at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:985)
```

Possible Causes

The user does not belong to the **kafkaadmin** group. Kafka provides a secure access interface. Only users in the **kafkaadmin** group can perform the setting operation.

Cause Analysis

1. After the client command is run, the "NoAuthException" exception is reported.
2. Run the client command **klist** to query the current authenticated user.

   ```
   [root@10-10-144-2 client]# klist
   Ticket cache: FILE:/tmp/krb5cc_0
   Default principal: test@HADOOP.COM

   Valid starting     Expires            Service principal
   01/25/17 11:06:48  01/26/17 11:06:45  krbtgt/HADOOP.COM@HADOOP.COM
   ```

   The **test** user is used in this example.
3. Run the **id** command to query the user group information.

   ```
   [root@10-10-144-2 client]# id test
   uid=20032(test) gid=10001(hadoop) groups=10001(hadoop),9998(ficommon),10003(kafka)
   ```

Solution

MRS Manager:

Step 1 Log in to MRS Manager.

Step 2 Choose **System** > **Manage User**.

Step 3 In the **Operation** column of the user, click **Modify**.

Step 4 Add the user to the **kafkaadmin** group.

Step 5 Run the **id** command to query the user group information.

```
[root@host1 client]# id test
uid=20032(test) gid=10001(hadoop) groups=10001(hadoop),9998(ficommon),10002(kafkaadmin),
10003(kafka)
```

----End

FusionInsight Manager:

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 Locate the row that contains the target user, and click **Modify**.

Step 4 Add the user to the **kafkaadmin** group. Click **OK**.

Step 5 Run the **id** command to query the user group information.

```
[root@10-10-144-2 client]# id test
uid=20032(test) gid=10001(hadoop) groups=10001(hadoop),9998(ficommon),10002(kafkaadmin),
10003(kafka)
```

----End

15.13.19 When a Kafka Topic Fails to Be Created, "NoNode for /brokers/ids" Is Displayed

Symptom

When running the following command on the Kafka client to create topics, it is found that the topics cannot be created.

```
kafka-topics.sh --create  --replication-factor 1 --partitions 2 --topic test --zookeeper
192.168.234.231:2181
```

The error message "NoNodeException: KeeperErrorCode = NoNode for /brokers/ids" is displayed.

Details are as follows:

```
Error while executing topic command : org.apache.zookeeper.KeeperException$NoNodeException:
KeeperErrorCode = NoNode for /brokers/ids
[2017-09-17 16:35:28,520] ERROR org.I0Itec.zkclient.exception.ZkNoNodeException:
org.apache.zookeeper.KeeperException$NoNodeException: KeeperErrorCode = NoNode for /brokers/ids
    at org.I0Itec.zkclient.exception.ZkException.create(ZkException.java:47)
    at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:995)
    at org.I0Itec.zkclient.ZkClient.getChildren(ZkClient.java:675)
    at org.I0Itec.zkclient.ZkClient.getChildren(ZkClient.java:671)
    at kafka.utils.ZkUtils.getChildren(ZkUtils.scala:541)
    at kafka.utils.ZkUtils.getSortedBrokerList(ZkUtils.scala:176)
    at kafka.admin.AdminUtils$.createTopic(AdminUtils.scala:235)
    at kafka.admin.TopicCommand$.createTopic(TopicCommand.scala:105)
    at kafka.admin.TopicCommand$.main(TopicCommand.scala:60)
    at kafka.admin.TopicCommand.main(TopicCommand.scala)
Caused by: org.apache.zookeeper.KeeperException$NoNodeException: KeeperErrorCode = NoNode for /
brokers/ids
    at org.apache.zookeeper.KeeperException.create(KeeperException.java:115)
    at org.apache.zookeeper.KeeperException.create(KeeperException.java:51)
    at org.apache.zookeeper.ZooKeeper.getChildren(ZooKeeper.java:2256)
    at org.apache.zookeeper.ZooKeeper.getChildren(ZooKeeper.java:2284)
    at org.I0Itec.zkclient.ZkConnection.getChildren(ZkConnection.java:114)
    at org.I0Itec.zkclient.ZkClient$4.call(ZkClient.java:678)
    at org.I0Itec.zkclient.ZkClient$4.call(ZkClient.java:675)
    at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:985)
    ... 8 more
 (kafka.admin.TopicCommand$)
```

Possible Causes

- The Kafka service is not running.
- The ZooKeeper address parameter in the client command is incorrectly configured.

Cause Analysis

1. After the client command is run, the "NoNodeException" exception is reported.

```
Error while executing topic command : org.apache.zookeeper.KeeperException$NoNodeException: 
KeeperErrorCode = NoNode for /brokers/ids
[2017-09-17 16:35:28,520] ERROR org.I0Itec.zkclient.exception.ZkNoNodeException: 
org.apache.zookeeper.KeeperException$NoNodeException: KeeperErrorCode = NoNode for /brokers/ids
    at org.I0Itec.zkclient.exception.ZkException.create(ZkException.java:47)
    at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:995)
    at org.I0Itec.zkclient.ZkClient.getChildren(ZkClient.java:675)
    at org.I0Itec.zkclient.ZkClient.getChildren(ZkClient.java:671)
    at kafka.utils.ZkUtils.getChildren(ZkUtils.scala:541)
    at kafka.utils.ZkUtils.getSortedBrokerList(ZkUtils.scala:176)
    at kafka.admin.AdminUtils$.createTopic(AdminUtils.scala:235)
    at kafka.admin.TopicCommand$.createTopic(TopicCommand.scala:105)
    at kafka.admin.TopicCommand$.main(TopicCommand.scala:60)
    at kafka.admin.TopicCommand.main(TopicCommand.scala)
```

2. Check whether the Kafka service is in the normal state on Manager.
3. Check whether the ZooKeeper address in the client command is correct. Check the Kafka information stored in ZooKeeper. The path (Znode) should be suffixed with **/kafka**. It is found that **/kafka** is missing in the configuration.

```
[root@10-10-144-2 client]#
kafka-topics.sh --create  --replication-factor 1 --partitions 2 --topic test --zookeeper 
192.168.234.231:2181
```

Solution

Step 1 Ensure that the Kafka service is normal.

Step 2 Add **/kafka** to the ZooKeeper address in the command.

```
[root@10-10-144-2 client]#
kafka-topics.sh --create  --replication-factor 1 --partitions 2 --topic test --zookeeper 
192.168.234.231:2181/kafka
```

----End

15.13.20 When a Kafka Topic Fails to Be Created, "replication factor larger than available brokers" Is Displayed

Symptom

When running the following command on the Kafka client to create topics, it is found that the topics cannot be created.

```
kafka-topics.sh --create  --replication-factor 2 --partitions 2 --topic test --zookeeper 
192.168.234.231:2181
```

The error message "replication factor larger than available brokers" is displayed.

See the following:

```
Error while executing topic command : replication factor: 2 larger than available brokers: 0
[2017-09-17 16:44:12,396] ERROR kafka.admin.AdminOperationException: replication factor: 2 larger than 
available brokers: 0
    at kafka.admin.AdminUtils$.assignReplicasToBrokers(AdminUtils.scala:117)
    at kafka.admin.AdminUtils$.createTopic(AdminUtils.scala:403)
    at kafka.admin.TopicCommand$.createTopic(TopicCommand.scala:110)
    at kafka.admin.TopicCommand$.main(TopicCommand.scala:61)
    at kafka.admin.TopicCommand.main(TopicCommand.scala)
 (kafka.admin.TopicCommand$)
```

Possible Causes

- The Kafka service is not running.

- The available Broker of the Kafka service is smaller than the configured **replication-factor**.
- The ZooKeeper address parameter in the client command is incorrectly configured.

Cause Analysis

1. After the client command is run, "replication factor larger than available brokers" is reported.
```
Error while executing topic command : replication factor: 2 larger than available brokers: 0
[2017-09-17 16:44:12,396] ERROR kafka.admin.AdminOperationException: replication factor: 2 larger
than available brokers: 0
    at kafka.admin.AdminUtils$.assignReplicasToBrokers(AdminUtils.scala:117)
    at kafka.admin.AdminUtils$.createTopic(AdminUtils.scala:403)
    at kafka.admin.TopicCommand$.createTopic(TopicCommand.scala:110)
    at kafka.admin.TopicCommand$.main(TopicCommand.scala:61)
    at kafka.admin.TopicCommand.main(TopicCommand.scala)
 (kafka.admin.TopicCommand$)
```
2. Check whether the Kafka service is in the normal state on Manager and whether the current available Broker is smaller than the configured **replication-factor**.
3. Check whether the ZooKeeper address in the client command is correct. Check the Kafka information stored in ZooKeeper. The path (Znode) should be suffixed with **/kafka**. It is found that **/kafka** is missing in the configuration.
```
[root@10-10-144-2 client]#
kafka-topics.sh --create  --replication-factor 2 --partitions 2 --topic test --zookeeper
192.168.234.231:2181
```

Solution

Step 1 Ensure that the Kafka service is in the normal state and the available Broker is not less than the configured **replication-factor**.

Step 2 Add **/kafka** to the ZooKeeper address in the command.
```
[root@10-10-144-2 client]#
kafka-topics.sh --create  --replication-factor 1 --partitions 2 --topic test --zookeeper
192.168.234.231:2181/kafka
```

----End

15.13.21 Consumer Repeatedly Consumes Data

Symptom

When the data volume is large, rebalance occurs frequently, causing repeated consumption. The key logs are as follows:

```
2018-05-12 10:58:42,561 | INFO | [kafka-request-handler-3] | [GroupCoordinator 2]: Preparing to restabilize
group DemoConsumer with old generation 118 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
2018-05-12 10:58:43,245 | INFO | [kafka-request-handler-5] | [GroupCoordinator 2]: Stabilized group
DemoConsumer generation 119 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
2018-05-12 10:58:43,560 | INFO | [kafka-request-handler-7] | [GroupCoordinator 2]: Assignment received
from leader for group DemoConsumer for generation 119 | kafka.coordinator.GroupCoordinator
(Logging.scala:68)
2018-05-12 10:59:13,562 | INFO | [executor-Heartbeat] | [GroupCoordinator 2]: Preparing to restabilize
group DemoConsumer with old generation 119 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
2018-05-12 10:59:13,790 | INFO | [kafka-request-handler-3] | [GroupCoordinator 2]: Stabilized group
DemoConsumer generation 120 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
2018-05-12 10:59:13,791 | INFO | [kafka-request-handler-0] | [GroupCoordinator 2]: Assignment received
```

```
from leader for group DemoConsumer for generation 120 | kafka.coordinator.GroupCoordinator
(Logging.scala:68)
2018-05-12 10:59:43,802 | INFO | [kafka-request-handler-2] | Rolled new log segment for
'__consumer_offsets-17' in 2 ms. | kafka.log.Log (Logging.scala:68)
2018-05-12 10:59:52,456 | INFO | [group-metadata-manager-0] | [Group Metadata Manager on Broker 2]:
Removed 0 expired offsets in 0 milliseconds. | kafka.coordinator.GroupMetadataManager (Logging.scala:68)
2018-05-12 11:00:49,772 | INFO | [kafka-scheduler-6] | Deleting segment 0 from log __consumer_offsets-17.
| kafka.log.Log (Logging.scala:68)
2018-05-12 11:00:49,773 | INFO | [kafka-scheduler-6] | Deleting index /srv/BigData/kafka/data4/kafka-logs/
__consumer_offsets-17/00000000000000000000.index.deleted | kafka.log.OffsetIndex (Logging.scala:68)
2018-05-12 11:00:49,773 | INFO | [kafka-scheduler-2] | Deleting segment 2147948547 from log
__consumer_offsets-17. | kafka.log.Log (Logging.scala:68)
2018-05-12 11:00:49,773 | INFO | [kafka-scheduler-4] | Deleting segment 4282404355 from log
__consumer_offsets-17. | kafka.log.Log (Logging.scala:68)
2018-05-12 11:00:49,775 | INFO | [kafka-scheduler-2] | Deleting index /srv/BigData/kafka/data4/kafka-logs/
__consumer_offsets-17/00000000002147948547.index.deleted | kafka.log.OffsetIndex (Logging.scala:68)
2018-05-12 11:00:49,775 | INFO | [kafka-scheduler-4] | Deleting index /srv/BigData/kafka/data4/kafka-logs/
__consumer_offsets-17/00000000004282404355.index.deleted | kafka.log.OffsetIndex (Logging.scala:68)
2018-05-12 11:00:50,533 | INFO | [kafka-scheduler-6] | Deleting segment 4283544095 from log
__consumer_offsets-17. | kafka.log.Log (Logging.scala:68)
2018-05-12 11:00:50,569 | INFO | [kafka-scheduler-6] | Deleting index /srv/BigData/kafka/data4/kafka-logs/
__consumer_offsets-17/00000000004283544095.index.deleted | kafka.log.OffsetIndex (Logging.scala:68)
2018-05-12 11:02:21,178 | INFO | [kafka-request-handler-2] | [GroupCoordinator 2]: Preparing to restabilize
group DemoConsumer with old generation 120 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
2018-05-12 11:02:22,839 | INFO | [kafka-request-handler-4] | [GroupCoordinator 2]: Stabilized group
DemoConsumer generation 121 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
2018-05-12 11:02:23,169 | INFO | [kafka-request-handler-1] | [GroupCoordinator 2]: Assignment received
from leader for group DemoConsumer for generation 121 | kafka.coordinator.GroupCoordinator
(Logging.scala:68)
2018-05-12 11:02:49,913 | INFO | [kafka-request-handler-6] | Rolled new log segment for
'__consumer_offsets-17' in 2 ms. | kafka.log.Log (Logging.scala:68)
```

In the logs, "Preparing to restabilize group DemoConsumer with old generation" indicates that rebalance occurs.

Possible Causes

The parameter settings are improper.

Cause Analysis

Cause: Due to improper parameter settings, the data processing time is too long when the data volume is large. Balance frequently occurs, and the offset cannot be submitted normally. As a result, the data is repeatedly consumed.

Principle: The offset is submitted only after the poll data is processed. If the processing duration after the poll data is processed exceeds the duration specified by **session.timeout.ms**, the rebalance occurs. As a result, the consumption fails and the offset of the consumed data cannot be submitted. Therefore, the data is consumed at the old offset next time. As a result, the data is repeatedly consumed.

Solution

Adjust the following service parameters on Manager:

request.timeout.ms=100000

session.timeout.ms=90000

max.poll.records=50

heartbeat.interval.ms=3000

Among the preceding parameters:

The value of **request.timeout.ms** is 10s greater than that of **session.timeout.ms**.

The value of **session.timeout.ms** must be within the values of **group.min.session.timeout.ms** and **group.max.session.timeout.ms** on the server.

Set the parameters as required. The **max.poll.records** parameter specifies the number of records for each poll. The purpose is to ensure that the processing time of poll data does not exceed the value of **session.timeout.ms**.

Related Information

- The post-poll data processing must be efficient and do not block the next poll.
- The poll method and data processing suggestion are processed asynchronously.

15.13.22 Leader for the Created Kafka Topic Partition Is Displayed as none

Symptom

When a user creates a topic using the Kafka client command, the leader for the created topic partition is displayed as **none**.

```
[root@10-10-144-2 client]#
kafka-topics.sh --create  --replication-factor 1 --partitions 2 --topic test --zookeeper 10.6.92.36:2181/
kafka

Created topic "test".
```

```
[root@10-10-144-2 client]#
kafka-topics.sh --describe --zookeeper 10.6.92.36:2181/kafka

Topic:test      PartitionCount:2        ReplicationFactor:2     Configs:
        Topic: test     Partition: 0    Leader: none       Replicas: 2,3   Isr:
        Topic: test     Partition: 1    Leader: none       Replicas: 3,1   Isr:
```

Possible Causes

- The Kafka service is not running.
- The user group information cannot be found.

Cause Analysis

1. Check the Kafka service status and monitoring metrics.
   - MRS Manager: Log in to MRS Manager and choose **Services** > **Kafka**. Check the Kafka status. The status is **Good**, and the monitoring metrics are correctly displayed.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Kafka**. Check the Kafka status. The status is **Good**, and the monitoring metrics are correctly displayed.
2. Obtain the Controller node information on the Kafka overview page.

3. Log in to the node where the Controller resides, and run the **cd /var/log/Bigdata/kafka/broker** command to go to the node log directory. The **state-change.log** contains "NoAuthException", which indicates that the ZooKeeper permission is incorrect.

```
2018-05-31 09:20:42,436 | ERROR | [ZkClient-
EventThread-34-10.6.92.36:24002,10.6.92.37:24002,10.6.92.38:24002/kafka] | Controller 4 epoch 6
initiated state change for partition [test,1] from NewPartition to OnlinePartition failed |
state.change.logger (Logging.scala:103)

 org.I0Itec.zkclient.exception.ZkException: org.apache.zookeeper.KeeperException$NoAuthException:
KeeperErrorCode = NoAuth for /brokers/topics/test/partitions
 at org.I0Itec.zkclient.exception.ZkException.create(ZkException.java:68)
 at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:1000)
 at org.I0Itec.zkclient.ZkClient.create(ZkClient.java:527)
 at org.I0Itec.zkclient.ZkClient.createPersistent(ZkClient.java:293)
```

4. Check on ZooKeeper audit logs recorded in the specified period also indicates that the permission is abnormal.

```
2018-05-31 09:20:42,421 | ERROR | CommitProcWorkThread-1 | session=0xc3000007015d5a18
user=10.6.92.39,kafka/hadoop.hadoop.com@HADOOP.COM,kafka/
hadoop.hadoop.com@HADOOP.COM    ip=10.6.92.39    operation=create znode
target=ZooKeeperServer    znode=/kafka/brokers/topics/test/partitions/0/state    result=failure
2018-05-31 09:20:42,423 | ERROR | CommitProcWorkThread-1 | session=0xc3000007015d5a18
user=10.6.92.39,kafka/hadoop.hadoop.com@HADOOP.COM,kafka/
hadoop.hadoop.com@HADOOP.COM    ip=10.6.92.39    operation=create znode
target=ZooKeeperServer    znode=/kafka/brokers/topics/test/partitions/0    result=failure
2018-05-31 09:20:42,435 | ERROR | CommitProcWorkThread-1 | session=0xc3000007015d5a18
user=10.6.92.39,kafka/hadoop.hadoop.com@HADOOP.COM,kafka/
hadoop.hadoop.com@HADOOP.COM    ip=10.6.92.39    operation=create znode
target=ZooKeeperServer    znode=/kafka/brokers/topics/test/partitions    result=failure
2018-05-31 09:20:42,439 | ERROR | CommitProcWorkThread-1 | session=0xc3000007015d5a18
user=10.6.92.39,kafka/hadoop.hadoop.com@HADOOP.COM,kafka/
hadoop.hadoop.com@HADOOP.COM    ip=10.6.92.39    operation=create znode
target=ZooKeeperServer    znode=/kafka/brokers/topics/test/partitions/1/state    result=failure
2018-05-31 09:20:42,441 | ERROR | CommitProcWorkThread-1 | session=0xc3000007015d5a18
user=10.6.92.39,kafka/hadoop.hadoop.com@HADOOP.COM,kafka/
hadoop.hadoop.com@HADOOP.COM    ip=10.6.92.39    operation=create znode
target=ZooKeeperServer    znode=/kafka/brokers/topics/test/partitions/1    result=failure
2018-05-31 09:20:42,453 | ERROR | CommitProcWorkThread-1 | session=0xc3000007015d5a18
user=10.6.92.39,kafka/hadoop.hadoop.com@HADOOP.COM,kafka/
hadoop.hadoop.com@HADOOP.COM    ip=10.6.92.39    operation=create znode
target=ZooKeeperServer    znode=/kafka/brokers/topics/test/partitions    result=failure
```

5. Run the **id -Gn kafka** command on each ZooKeeper instance node. It is found that user group information cannot be queried on a node.

```
[root @bdpsit3ap03 ~]# id -Gn kafka
id:  kafka: No such user
[root @bdpsit3ap03 ~]#
```

6. In an MRS cluster, user management is provided by the LDAP service and depends on the SSSD (Red Hat) and NSCD (SUSE) services of OSs. The process from creating a user to synchronizing the user to the SSSD service takes some time. If the user does not take effect or the SSSD version has bugs, the user may be invalid on the ZooKeeper node in some cases, which causes topic creation failures.

Solution

Step 1 Restart the SSD/NSCD service.

- RedHat

```
service sssd restart
```

- SUSE

```
sevice nscd restart
```

Step 2 After restarting related services, run the **id username** command on the active ResourceManager node to check whether the user information is valid.

----End

15.13.23 Safety Instructions on Using Kafka

Brief Introduction to API for Kafka

- New Producer API

  Indicates the API defined in org.apache.kafka.clients.producer.KafkaProducer. When kafka-console-producer.sh is used, the API is used by default.
- Old Producer API

  Indicates the API defined in kafka.producer.Producer. When kafka-console-producer.sh is used, the API is invoked to add --old-producer.
- New Consumer API

  Indicates the API defined in org.apache.kafka.clients.consumer.KafkaConsumer. When kafka-console-consumer.sh is used, the API is invoked to add --new-consumer.
- Old Consumer API

  Indicates the API defined in kafka.consumer.ConsumerConnector. When **kafka-console-consumer.sh** is used, the API is used by default.

NOTE

New Producer API and new Consumer API are called new API in general in the document.

Protocol Description for Accessing Kafka

The protocols used to access Kafka are as follows: PLAINTEXT, SSL, SASL_PLAINTEXT, and SASL_SSL.

When Kafka service is started, the listeners using the PLAINTEXT and SASL_PLAINTEXT protocols are started. You can set **ssl.mode.enable** to **true** in Kafka service configuration to start listeners using SSL and SASL_SSL protocols.

The following table describes the four protocols:

Protocol Type	Description	Supported API	Default Port
PLAINTEXT	Supports plaintext access without authentication.	New and old APIs	9092
SASL_PLAINTEXT	Supports plaintext access with Kerberos authentication.	New API	21007
SSL	Supports SSL-encrypted access without authentication.	New API	9093
SASL_SSL	Supports SSL-encrypted access with Kerberos authentication.	New API	21009

ACL Settings for Topic

Kafka supports secure access. Therefore, users can set the ACL for topics to control that different users access different topics. To view and set the permission information about a topic, run the kafka-acls.sh script on the Linux client.

- Scenarios

  Assign Kafka users with specific permissions for related topics based on service requirements.

  The following table describes default Kafka user groups.

User Group	Description
kafkaadmin	Kafka administrator group. Users added to this group have the permissions to create, delete, authorize, as well as read from and write data to all topics.
kafkasuperuser	Users added to this group have permissions to read data from and write data to all topics.
kafka	Kafka common user group. If users in this group want to read data from and write data to a specific topic, the users in the kafkaadmin group must grant permissions to users in this group.

- Prerequisites
  1. The system administrator has understood service requirements and prepared a Kafka administrator (belonging to the kafkaadmin group).
  2. The Kafka client has been installed.
- Procedure
  1. Log in to the node where the Kafka client is installed as the client installation user.
  2. Switch to the Kafka client installation directory, for example, **/opt/kafkaclient**.

     cd /opt/kafkaclient
  3. Run the following command to configure environment variables:

     source bigdata_env
  4. Run the following command to perform user authentication (skip this step for a cluster in common mode):

     kinit *Component service user*
  5. Run the following command to switch to the Kafka client installation directory:

     cd Kafka/kafka/bin
  6. The following describes the commands commonly used for user authorization when **kafka-acl.sh** is used:

- View the permission control list of a topic:

  ./kafka-acls.sh --authorizer-properties zookeeper.connect=<ZooKeeper cluster service IP:2181/kafka > --list --topic <Topic name>

- Add the Producer permission for a user:

  ./kafka-acls.sh --authorizer-properties zookeeper.connect=<ZooKeeper cluster service IP:2181/kafka > --add --allow-principal User:<username> --producer --topic <Topic name>

- Remove the Producer permission from a user:

  ./kafka-acls.sh --authorizer-properties zookeeper.connect=<ZooKeeper cluster service IP:2181/kafka > --remove --allow-principal User:<username> --producer --topic <Topic name>

- Add the Consumer permission for a user:

  ./kafka-acls.sh --authorizer-properties zookeeper.connect=<ZooKeeper cluster service IP:2181/kafka > --add --allow-principal User:<username> --consumer --topic <Topic name> --group <consumer group name>

- Remove the Consumer permission from a user:

  ./kafka-acls.sh --authorizer-properties zookeeper.connect=<ZooKeeper cluster service IP:2181/kafka > --remove --allow-principal User:<username> --consumer --topic <Topic name> --group <consumer group name>

Use of New and Old Kafka APIs in Different Scenarios

- Scenario 1: accessing the topic with an ACL

Used API	User Group	Client Parameter	Server Parameter	Access Port
New API	Users need to meet one of the following conditions: ● In the administrator group ● In the **kafkaadmin** group ● In the **kafkasup**	security.protocol=SASL_PLAINTEXT sasl.kerberos.service.name = kafka	-	sasl.port (The default number is 21007.)

Used API	User Group	Client Parameter	Server Parameter	Access Port

	eruser group
	● In the **kafka** group and be authorized
		security.protocol=SASL_SSL
		sasl.kerberos.service.name = kafka
			Set ssl.mode.enable to true.	sasl-ssl.port (The default port number is 21009.)

Old API	N/A	N/A	N/A	N/A

- Scenario 2: accessing the topic without an ACL

Used API	User Group	Client Parameter	Server Parameter	Access Port

New API	Users need to meet one of the following conditions:
	● In the administrator group
	● In the **kafkaadmin** group
	● In the **kafkasuperuser** group
		security.protocol=SASL_PLAINTEXT
		sasl.kerberos.service.name = kafka
			-	sasl.port (The default number is 21007.)

	Users are in the **kafka** group.		Set **allow.everyone.if.no.acl.found** to **true**.	sasl.port (The default number is 21007.)

Used API	User Group	Client Parameter	Server Parameter	Access Port
	Users need to meet one of the following conditions: ● In the administrator group ● In the **kafkaadmin** group ● In the **kafkasuperuser** group	security.protocol=SASL_SSLsasl.kerberos.service.name = kafka	Set **ssl-enable** to **true**.	sasl-ssl.port (The default port number is 21009.)
	Users are in the **kafka** group.		Set **allow.everyone.if.no.acl.found** to **true**. Set **ssl-enable** to **true**.	sasl-ssl.port (The default port number is 21009.)
	-	security.protocol=PLAINTEXT	Set **allow.everyone.if.no.acl.found** to **true**.	port (The default number is 21005.)
	-	security.protocol=SSL	Set **allow.everyone.if.no.acl.found** to **true**. Set **ssl-enable** to **true**.	ssl.port (The default number is 21008.)
Old Producer	-	-	Set **allow.everyone.if.no.acl.found** to **true**.	port (The default number is 21005.)
Old Consumer	-	-	Set **allow.everyone.if.no.acl.found** to **true**.	ZooKeeper service port: clientPort (The default number is 24002.)

15.13.24 Obtaining Kafka Consumer Offset Information

Symptom

How do I obtain Kafka Consumer offset information when using Kafka Consumer to consume data?

Kafka APIs

- New Producer API

  Indicates the API defined in **org.apache.kafka.clients.producer.KafkaProducer**. When **kafka-console-producer.sh** is used, the API is used by default.
- Old Producer API

  Indicates the API defined in **kafka.producer.Producer**. When **kafka-console-producer.sh** is used, the API is invoked to add **--old-producer**.
- New Consumer API

  Indicates the API defined in **org.apache.kafka.clients.consumer.KafkaConsumer**. When **kafka-console-consumer.sh** is used, the API is invoked to add **--new-consumer**.
- Old Consumer API

  Indicates the API defined in **kafka.consumer.ConsumerConnector**. When **kafka-console-consumer.sh** is used, the API is used by default.

NOTE

New Producer API and new Consumer API are called new API in general in the document.

Procedure

Old Consumer API

- Prerequisites
  1. a. The system administrator has understood service requirements and prepared a Kafka administrator (belonging to the kafkaadmin group).
  2. b. The Kafka client has been installed.
- Procedure
  1. a. Log in to the node where the Kafka client is installed as the client installation user.
  2. b. Switch to the Kafka client installation directory, for example, **/opt/kafkaclient**.

     cd /opt/kafkaclient
  3. c. Run the following command to configure environment variables:

     source bigdata_env
  4. d. Run the following command to perform user authentication (skip this step for a cluster in common mode):

     kinit *Component service user*

e. Run the following command to switch to the Kafka client installation directory:

   cd Kafka/kafka/bin

f. Run the following command to obtain Consumer offset metric information:

```
 bin/kafka-consumer-groups.sh --zookeeper <zookeeper_host:port>/kafka --list
```

```
bin/kafka-consumer-groups.sh --zookeeper <zookeeper_host:port>/kafka --describe --group
test-consumer-group
```

   Example:

```
kafka-consumer-groups.sh --zookeeper 192.168.100.100:2181/kafka --list
kafka-consumer-groups.sh --zookeeper 192.168.100.100:2181/kafka --describe --group test-
consumer-group
```

New Consumer API

- Prerequisites
  a. The system administrator has understood service requirements and prepared a Kafka administrator (belonging to the kafkaadmin group).
  b. The Kafka client has been installed.
- Procedure
  a. Log in to the node where the Kafka client is installed as the client installation user.
  b. Switch to the Kafka client installation directory, for example, **/opt/client**.

     cd /opt/client

  c. Run the following command to configure environment variables:

     source bigdata_env

  d. Run the following command to perform user authentication (skip this step for a cluster in common mode):

     kinit *Component service user*

  e. Run the following command to switch to the Kafka client installation directory:

     cd Kafka/kafka/bin

  f. Run the following command to obtain Consumer offset metric information:

     kafka-consumer-groups.sh --bootstrap-server <broker_host:port> --describe --group my-group

     Example:

     kafka-consumer-groups.sh --bootstrap-server 192.168.100.100:9092 --describe --group my-group

15.13.25 Adding or Deleting Configurations for a Topic

Symptom

Configure or modify a specific topic when using Kafka.

Parameters that can be modified at the topic level:

```
cleanup.policy
compression.type
delete.retention.ms
file.delete.delay.ms
flush.messages
flush.ms
index.interval.bytes
max.message.bytes
min.cleanable.dirty.ratio
min.insync.replicas
preallocate
retention.bytes
retention.ms
segment.bytes
segment.index.bytes
segment.jitter.ms
segment.ms
unclean.leader.election.enable
```

Procedure

- Prerequisites

  The Kafka client has been installed.

- Procedure

  a. Log in to the node where the Kafka client is installed as the client installation user.

  b. Switch to the Kafka client installation directory, for example, **/opt/client**.

     cd /opt/client

  c. Run the following command to configure environment variables:

     source bigdata_env

  d. Run the following command to perform user authentication (skip this step for a cluster in common mode):

     kinit *Component service user*

  e. Run the following command to switch to the Kafka client installation directory:

     cd Kafka/kafka/bin

  f. Run the following commands to configure and delete a topic:

     kafka-topics.sh --alter --topic <topic_name> --zookeeper <zookeeper_host:port>/kafka --config <name=value>

     kafka-topics.sh --alter --topic <topic_name> --zookeeper <zookeeper_host:port>/kafka --delete-config <name>

     Example:

     kafka-topics.sh --alter --topic test1 --zookeeper 192.168.100.100:2181/kafka --config retention.ms=86400000

     kafka-topics.sh --alter --topic test1 --zookeeper 192.168.100.100:2181/kafka --delete-config retention.ms

  g. Run the following command to query topic information:

     kafka-topics.sh --describe -topic <topic_name> --zookeeper <zookeeper_host:port>/kafka

15.13.26 Reading the Content of the __consumer_offsets Internal Topic

Issue

How does Kafka save the offset of a Consumer to the **__consumer_offsets** of internal topics?

Procedure

Step 1 Log in to the node where the Kafka client is installed as the client installation user.

Step 2 Switch to the Kafka client installation directory, for example, **/opt/client**.

cd /opt/client

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 Run the following command to perform user authentication (skip this step for a cluster in common mode):

kinit *Component service user*

Step 5 Run the following command to switch to the Kafka client installation directory:

cd Kafka/kafka/bin

Step 6 Run the following command to obtain Consumer offset metric information:

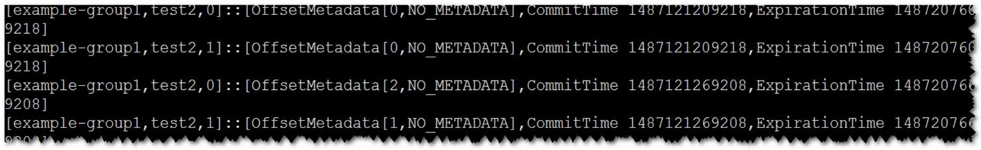
kafka-console-consumer.sh --topic __consumer_offsets --zookeeper <zk_host:port>/kafka --formatter "kafka.coordinator.group.GroupMetadataManager\$OffsetsMessageFormatter" --consumer.config <property file> --from-beginning

Add the following content to the <*property file*> configuration file:

```
exclude.internal.topics = false
```

Example:

kafka-console-consumer.sh --topic __consumer_offsets --zookeeper 10.5.144.2:2181/kafka --formatter "kafka.coordinator.group.GroupMetadataManager\$OffsetsMessageFormatter" --consumer.config ../config/consumer.properties --from-beginning



```
[example-group1,test2,0]::[OffsetMetadata[0,NO_METADATA],CommitTime 1487121209218,ExpirationTime 14872076
9218]
[example-group1,test2,1]::[OffsetMetadata[0,NO_METADATA],CommitTime 1487121209218,ExpirationTime 14872076
9218]
[example-group1,test2,0]::[OffsetMetadata[2,NO_METADATA],CommitTime 1487121269208,ExpirationTime 14872076
9208]
[example-group1,test2,1]::[OffsetMetadata[1,NO_METADATA],CommitTime 1487121269208,ExpirationTime 14872076
```

----End

15.13.27 Configuring Logs for Shell Commands on the Client

Issue

How do I set the log level for shell commands on the client?

Procedure

Step 1 Log in to the node where the Kafka client is installed as the client installation user.

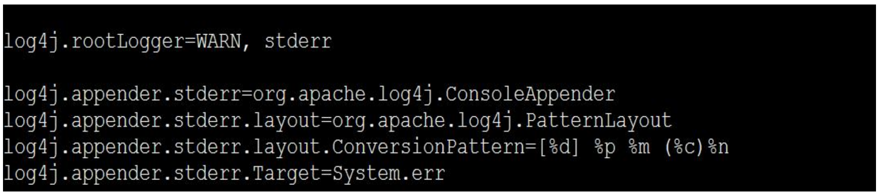
Step 2 Switch to the Kafka client installation directory, for example, **/opt/client**.

cd /opt/client

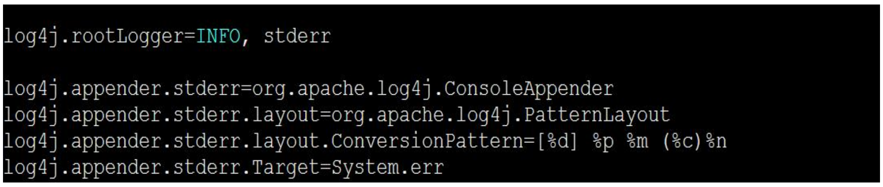
Step 3 Run the following command to switch to the Kafka client configuration directory:

cd Kafka/kafka/config

Step 4 Open the **tools-log4j.properties** file, change **WARN** to **INFO**, and save the file.



```
log4j.rootLogger=WARN, stderr

log4j.appender.stderr=org.apache.log4j.ConsoleAppender
log4j.appender.stderr.layout=org.apache.log4j.PatternLayout
log4j.appender.stderr.layout.ConversionPattern=[%d] %p %m (%c)%n
log4j.appender.stderr.Target=System.err
```



```
log4j.rootLogger=INFO, stderr

log4j.appender.stderr=org.apache.log4j.ConsoleAppender
log4j.appender.stderr.layout=org.apache.log4j.PatternLayout
log4j.appender.stderr.layout.ConversionPattern=[%d] %p %m (%c)%n
log4j.appender.stderr.Target=System.err
```

Step 5 Switch to the Kafka client installation directory, for example, **/opt/client**.

cd /opt/client

Step 6 Run the following command to configure environment variables:

source bigdata_env

Step 7 Run the following command to perform user authentication (skip this step for a cluster in common mode):

kinit *Component service user*

Step 8 Run the following command to switch to the Kafka client installation directory:

cd Kafka/kafka/bin

Step 9 Run the following command to obtain the topic information. The log information can be viewed on the console.

```
kafka-topics.sh --list --zookeeper 10.5.144.2:2181/kafka
[2017-02-17 14:34:27,005] INFO JAAS File name: /opt/client/Kafka/./kafka/config/jaas.conf
(org.I0Itec.zkclient.ZkClient)
[2017-02-17 14:34:27,007] INFO Starting ZkClient event thread. (org.I0Itec.zkclient.ZkEventThread)
[2017-02-17 14:34:27,013] INFO Client environment:zookeeper.version=V100R002C10, built on 05/12/2016
08:56 GMT (org.apache.zookeeper.ZooKeeper)
[2017-02-17 14:34:27,013] INFO Client environment:host.name=10-10-144-2
(org.apache.zookeeper.ZooKeeper)
```

```
[2017-02-17 14:34:27,013] INFO Client environment:java.version=1.8.0_72
(org.apache.zookeeper.ZooKeeper)
[2017-02-17 14:34:27,013] INFO Client environment:java.vendor=Oracle Corporation
(org.apache.zookeeper.ZooKeeper)
[2017-02-17 14:34:27,013] INFO Client environment:java.home=/opt/client/JDK/jdk/jre
(org.apache.zookeeper.ZooKeeper)
Test
__consumer_offsets
counter
test
test2
test3
test4
```

----End

15.13.28 Obtaining Topic Distribution Information

Issue

How do I obtain topic distribution information in a Broker instance?

Preparations

- Prerequisites

  The Kafka and ZooKeeper clients have been installed.

- Procedure

  a. Log in to the node where the Kafka client is installed as the client installation user.

  b. Switch to the Kafka client installation directory, for example, **/opt/client**.

     cd /opt/client

  c. Run the following command to configure environment variables:

     source bigdata_env

  d. Run the following command to perform user authentication (skip this step for a cluster in common mode):

     kinit *Component service user*

  e. Run the following command to switch to the Kafka client installation directory:

     cd Kafka/kafka/bin

  f. Run the Kafka commands to obtain the topic assignment information and copy synchronization information, and check the return result.

     kafka-topics.sh --describe --zookeeper <zk_host:port/chroot>

     Example:

     ```
     [root@mgtdat-sh-3-01-3 client]#kafka-topics.sh --describe --zookeeper 10.149.0.90:2181/
     kafka
     Topic:topic1    PartitionCount:2   ReplicationFactor:2     Configs:
     Topic: topic1   Partition: 0 Leader: 26 Replicas: 23,25 Isr: 26
     Topic: topic1   Partition: 1 Leader: 24 Replicas: 24,23 Isr: 24,23
     ```

     In the preceding information, **Replicas** indicates the replica assignment information and **Isr** indicates the replica synchronization information.

Solution 1

1. Query the Broker ID mapping in ZooKeeper.

   sh zkCli.sh -server <zk_host:port>

2. Run the following command on the ZooKeeper client:

   ls /kafka/brokers/ids

   get/kafka/brokers/ids/<queried Broker ID>

   Example:

```
[root@node-master1gAMQ kafka]# zkCli.sh -server node-master1gAMQ:2181
Connecting to node-master1gAMQ:2181
Welcome to ZooKeeper!
JLine support is enabled

WATCHER::

WatchedEvent state:SyncConnected type:None path:null
[zk: node-master1gAMQ:2181(CONNECTED) 0] ls /kafka/brokers/
ids      seqid    topics
[zk: node-master1gAMQ:2181(CONNECTED) 0] ls /kafka/brokers/ids
[1]
[zk: node-master1gAMQ:2181(CONNECTED) 1] get /kafka/brokers/ids/1
{"listener_security_protocol_map":{"PLAINTEXT":"PLAINTEXT","SSL":"SSL"},"endpoints":["PLAINTEXT://
192.168.2.242:9092","SSL://192.168.2.242:9093"],"rack":"/default/rack0","jmx_port":
21006,"host":"192.168.2.242","timestamp":"1580886124398","port":9092,"version":4}
[zk: node-master1gAMQ:2181(CONNECTED) 2]
```

Solution 2

Obtain the mapping between nodes and Broker IDs.

kafka-broker-info.sh --zookeeper <zk_host:port/chroot>

Example:

```
[root@node-master1gAMQ kafka]# bin/kafka-broker-info.sh --zookeeper 192.168.2.70:2181/kafka
Broker_ID     IP_Address
--------------------------
 1           192.168.2.242
```

15.13.29 Kafka HA Usage Description

Kafka High Reliability and Availability

Kafka message transmission assurance mechanism ensures message transmission after required parameters are set to meet different performance and reliability requirements.

- **Kafka high availability and high performance**

  If HA and high performance are required, configure parameters listed in the following table.

Parameter	Default Value	Description
unclean.leader.election.enable	true	Specifies whether a replica that is not in the ISR can be selected as the leader. If this parameter is set to **true**, data may be lost.
auto.leader.rebalance.enable	true	Specifies whether the leader automated balancing function is used. If this parameter is set to **true**, the controller periodically balances the leader of each partition on all nodes and assigns the leader to a replica with a higher priority.
acks	1	The leader needs to check whether the message has been received and determine whether the required operation has been processed. This parameter affects message reliability and performance. ● If this parameter is set to **0**, the Producer does not wait for any response from the server and the message is considered successful. ● If this parameter is set to **1**, when the leader of the copy verifies that data has been written into the cluster, the leader makes repose quickly without waiting until all the copies are written. In this case, if the leader is abnormal when the leader makes the confirmation but replica synchronization is not complete, data will be lost. ● If this parameter is set to **-1** (all), the synchronization is successful only after all synchronization copies are confirmed. If **min.insync.replicas** is also configured, multiple copies can be written successfully. In this case, as long as one copy remains active, the record is not lost. **NOTE** This parameter is configured in the Kafka client configuration file.
min.insync.replicas	1	Specifies the minimum number of replicas to which data is written when **acks** is set to **-1** for the Producer.

Impact of HA and high performance configurations:

NOTICE

After HA and high performance are configured, the data reliability decreases. Specifically, data may be lost of disks or nodes are faulty.

- **Kafka high reliability configuration**

  If high data reliability is required, configure parameters listed in the following table.

Parameter	Recommended Value	Description
unclean.leader.election.enable	false	Indicates whether a replica that is not in the ISR list can be elected as a leader.
acks	-1	The leader needs to check whether the message has been received and determine whether the required operation has been processed. If this parameter is set to **-1**, the message is successfully received only when all replicas in the ISR list have confirmed to receive the message. The **min.insync.replicas** parameter must also be set to ensure that multiple copies can be written successfully. As long as one copy is active, the record is not lost. **NOTE** This parameter is configured in the Kafka client configuration file.
min.insync.replicas	2	Specifies the minimum number of replicas to which data is written when **acks** is set to **-1** for the Producer. Ensure that the value of **Min.insync.replicas** is equal to or less than that of **replication.factor**.

Impact of high reliability configurations:

– Deteriorated performance

  All copies in the ISR list are required, and the writing of the minimum number of copies has been verified successful. As a result, the delay of a single message increases and the processing capability of the client decreases. The actual performance depends on the onsite test data.

– Reduced availability

  A replica that is not in the ISR list cannot be elected as a leader. If the leader goes offline and other replicas are not in the ISR list, the partition remains unavailable until the leader node recovers.

  All copies in the ISR list are required, and the writing of the minimum number of copies has been verified successful. When the node where a copy of a partition is located is faulty, the minimum number of successful copies cannot be met. As a result, service writing fails.

Configuration Impact

Evaluate reliability and performance requirements based on service scenarios and use proper parameter configuration.

NOTE

- For valuable data, you are advised to configure raid1 or raid5 for Kafka data directory disks to improve data reliability in case disk fault of a single disk.
- The **acks** parameter is named different for different Producer APIs.
  - New Producer API

    Indicates the interface defined in **org.apache.kafka.clients.producer.KafkaProducer**. The **acks** parameter name remains unchanged for this API.
  - Old Producer API

    Indicates the interface defined in **kafka.producer.Producer**. The **acks** parameter is named as **request.required.acks** for this API.
- For parameters that can be modified at the topic level, the service level configurations are used by default. These parameters can be separately configured based on topic reliability requirements.

  For example, you can configure the reliability parameters of the topic named **test**.

  kafka-topics.sh --zookeeper 192.168.1.205:2181/kafka --alter --topic test --config unclean.leader.election.enable=false --config min.insync.replicas=2

  192.168.1.205 indicates the ZooKeeper service IP address.
- If modification of the service-level requires the restart of Kafka, you are advised to modify the service-level configuration on the change page.

15.13.30 Kafka Producer Writes Oversized Records

Symptom

When a user develops a Kafka application and invokes the new interface (**org.apache.kafka.clients.producer.***) as a Producer to write data to Kafka, the size of a single record is 1100055, which exceeds the value (**1000012**) of **message.max.bytes** in the Kafka configuration file **server.properties**. After the values of **message.max.bytes** and **replica.fetch.max.bytes** in the Kafka service configuration are changed to **5242880**, the exception persists. The error information is as follows:

```
..........
14749 [Thread-0] INFO  com.XXXXXX.bigdata.kafka.example.NewProducer  - The ExecutionException 
occured : {}.
java.util.concurrent.ExecutionException: org.apache.kafka.common.errors.RecordTooLargeException: The 
message is 1100093 bytes when serialized which is larger than the maximum request size you have 
configured with the max.request.size configuration.
at org.apache.kafka.clients.producer.KafkaProducer$FutureFailure.<init>(KafkaProducer.java:739)
at org.apache.kafka.clients.producer.KafkaProducer.doSend(KafkaProducer.java:483)
at org.apache.kafka.clients.producer.KafkaProducer.send(KafkaProducer.java:430)
at org.apache.kafka.clients.producer.KafkaProducer.send(KafkaProducer.java:353)
at com.XXXXXX.bigdata.kafka.example.NewProducer.run(NewProducer.java:150)
Caused by: org.apache.kafka.common.errors.RecordTooLargeException: The message is **** bytes when 
serialized which is larger than the maximum request size you have configured with the max.request.size 
configuration.
.......
```

Cause Analysis

When data is written to Kafka, the Kafka client compares the value of **max.request.size** with the size of the data to be written. If the size of the data to be written exceeds the default value of **max.request.size**, the preceding exception is reported.

Solution

Step 1 You can set the value of **max.request.size** when initializing the Kafka Producer instance.

For example, you can set this parameter to **5252880** as follows:

```
    // Protocol type: Currently, the SASL_PLAINTEXT or PLAINTEXT protocol types can be used.
    props.put(securityProtocol, kafkaProc.getValues(securityProtocol, "SASL_PLAINTEXT"));
     // Service name
    props.put(saslKerberosServiceName, "kafka");
    props.put("max.request.size", "5252880");
    .......
```

----End

15.13.31 Kafka Consumer Reads Oversized Records

Symptom

After data is written to Kafka, a user develops an application and invokes the interface (**org.apache.kafka.clients.consumer.***) to read data from Kafka as a Consumer. However, the reading fails and the following error is reported:

```
..........
1687 [KafkaConsumerExample] INFO  org.apache.kafka.clients.consumer.internals.AbstractCoordinator  - 
Successfully joined group DemoConsumer with generation 1
1688 [KafkaConsumerExample] INFO  org.apache.kafka.clients.consumer.internals.ConsumerCoordinator  - 
Setting newly assigned partitions [default-0, default-1, default-2] for group DemoConsumer
2053 [KafkaConsumerExample] ERROR com.XXXXXX.bigdata.kafka.example.NewConsumer  - 
[KafkaConsumerExample], Error due to 
org.apache.kafka.common.errors.RecordTooLargeException: There are some messages at [Partition=Offset]: 
{default-0=177} whose size is larger than the fetch size 1048576 and hence cannot be ever returned. 
Increase the fetch size on the client (using max.partition.fetch.bytes), or decrease the maximum message 
size the broker will allow (using message.max.bytes).
2059 [KafkaConsumerExample] INFO  com.XXXXXX.bigdata.kafka.example.NewConsumer  - 
[KafkaConsumerExample], Stopped
.......
```

Cause Analysis

When reading data, the Kafka client compares the size of the data to be read with the value of **max.partition.fetch.bytes**. If the size exceeds the value of **max.partition.fetch.bytes**, the preceding exception is reported.

Solution

Step 1 When creating a Kafka Consumer instance during initialization, set **max.partition.fetch.bytes**.

For example, you can set this parameter to **5252880** as follows:

```
......
// Security protocol type
```

```
props.put(securityProtocol, kafkaProc.getValues(securityProtocol, "SASL_PLAINTEXT"));
// Service name
props.put(saslKerberosServiceName, "kafka");

props.put("max.partition.fetch.bytes","5252880");
......
```

----End

15.13.32 High Usage of Multiple Disks on a Kafka Cluster Node

Issue

The usage of multiple disks on a node in the Kafka streaming cluster is high. The Kafka service will become unavailable if the usage reaches 100%.

Symptom

A node in the MRS Kafka streaming cluster created by the customer has multiple disks. Due to improper partitioning and service reasons, the usage of some disks is high. When the usage reaches 100%, Kafka becomes unavailable.

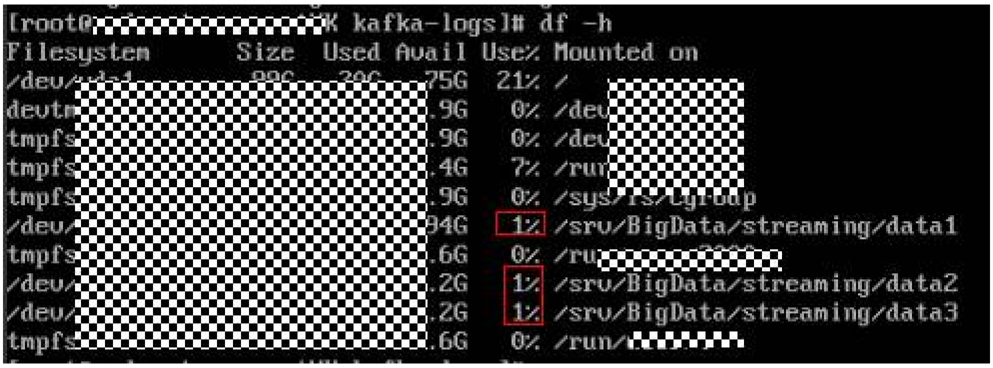
Cause Analysis

The disk data needs to be processed in a timely manner. After the value of **log.retention.hours** is changed, the service needs to be restarted. To ensure service continuity, you can shorten the aging time of a single data-intensive topic as required.

Procedure

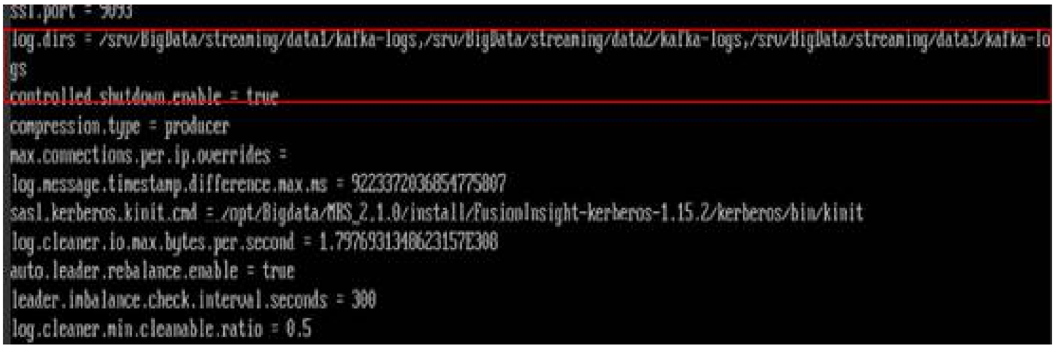
Step 1 Log in to the core node of the Kafka streaming cluster.

Step 2 Run the **df –h** command to check the disk usage.



```
[root@[redacted]K kafka-logs]# df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/[redacted]         75G  21% /
devtm[redacted]         .9G   0% /dev[redacted]
tmpfs[redacted]         .9G   0% /dev
tmpfs[redacted]         .4G   7% /run
tmpfs[redacted]         .9G   0% /sys/fs/cgroup
/dev/[redacted]         94G   1% /srv/BigData/streaming/data1
tmpfs[redacted]         .6G   0% /ru[redacted]
/dev/[redacted]         .2G   1% /srv/BigData/streaming/data2
/dev/[redacted]         .2G   1% /srv/BigData/streaming/data3
tmpfs[redacted]         .6G   0% /run/[redacted]
```

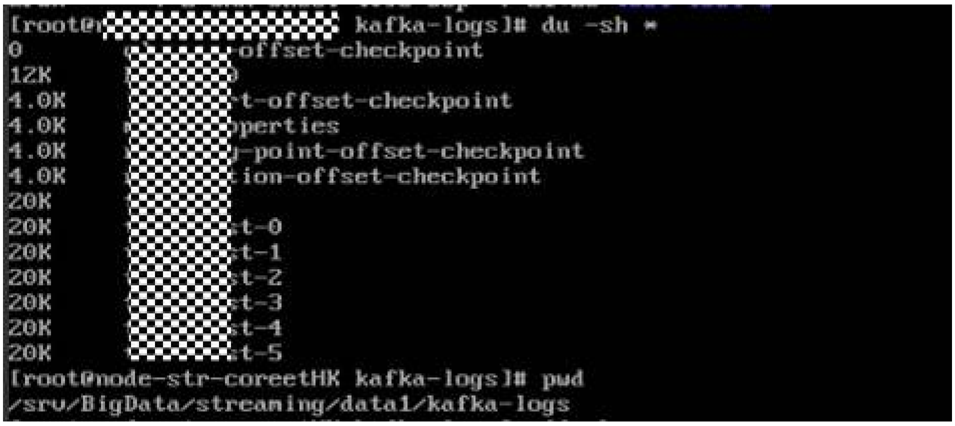
Step 3 Obtain the data storage directory from the **log.dirs** configuration item in the Kafka configuration file **opt/Bigdata/MRS_2.1.0/1_11_Broker/etc/server.properties**. Change the configuration file path based on the cluster version in the environment. If there are multiple disks, use commas (,) to separate multiple configuration items.



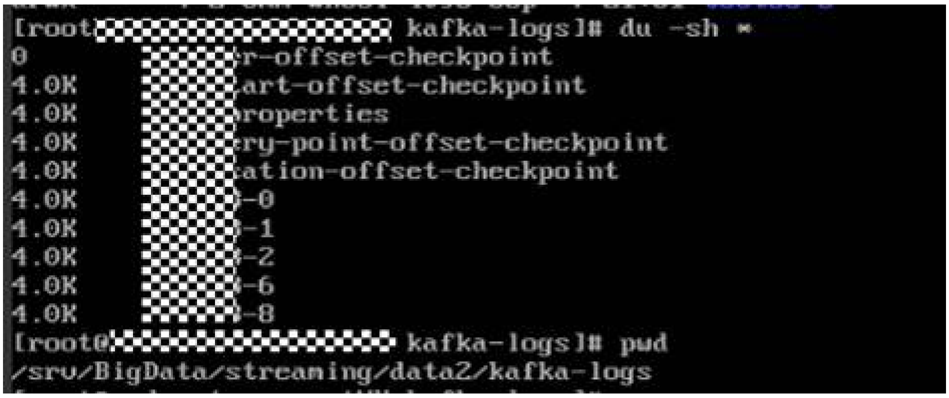
```
ssl.port = 9093
log.dirs = /srv/BigData/streaming/data1/kafka-logs,/srv/BigData/streaming/data2/kafka-logs,/srv/BigData/streaming/data3/kafka-lo
gs
controlled.shutdown.enable = true
compression.type = producer
max.connections.per.ip.overrides =
log.message.timestamp.difference.max.ms = 9223372036854775807
sasl.kerberos.kinit.cmd = /opt/Bigdata/MRS_2.1.0/install/FusionInsight-kerberos-1.15.2/kerberos/bin/kinit
log.cleaner.io.max.bytes.per.second = 1.7976931348623157E308
auto.leader.rebalance.enable = true
leader.imbalance.check.interval.seconds = 300
log.cleaner.min.cleanable.ratio = 0.5
```

Step 4 Run the **cd** command to go to the data storage directory obtained in **Step 3** of the disk with high usage.

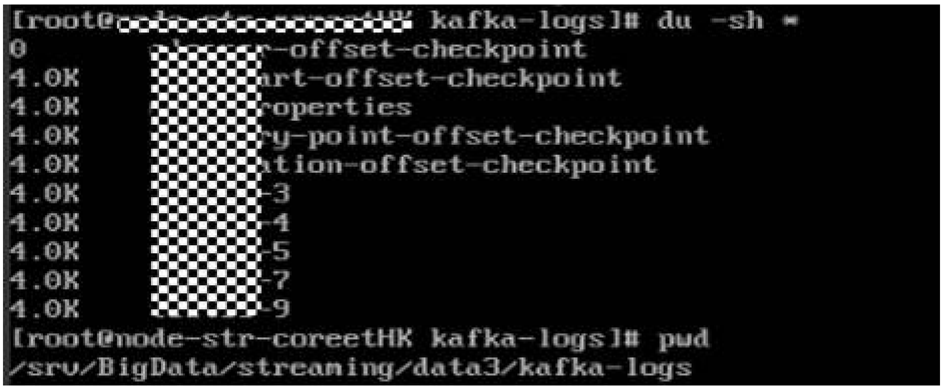
Step 5 Run the **du -sh *** command to print the name and size of the current topic.



```
[root@[redacted] kafka-logs]# du -sh *
0       [redacted]offset-checkpoint
12K     [redacted]
4.0K    [redacted]t-offset-checkpoint
4.0K    [redacted]perties
4.0K    [redacted]-point-offset-checkpoint
4.0K    [redacted]ion-offset-checkpoint
20K     [redacted]
20K     [redacted]t-0
20K     [redacted]t-1
20K     [redacted]t-2
20K     [redacted]t-3
20K     [redacted]t-4
20K     [redacted]t-5
[root@node-str-coreetHK kafka-logs]# pwd
/srv/BigData/streaming/data1/kafka-logs
```



```
[root[redacted] kafka-logs]# du -sh *
0       [redacted]r-offset-checkpoint
4.0K    [redacted]art-offset-checkpoint
4.0K    [redacted]roperties
4.0K    [redacted]ry-point-offset-checkpoint
4.0K    [redacted]ation-offset-checkpoint
4.0K    [redacted]-0
4.0K    [redacted]-1
4.0K    [redacted]-2
4.0K    [redacted]-6
4.0K    [redacted]-8
[root@[redacted] kafka-logs]# pwd
/srv/BigData/streaming/data2/kafka-logs
```



```
[root@[redacted] kafka-logs]# du -sh *
0       [redacted]r-offset-checkpoint
4.0K    [redacted]art-offset-checkpoint
4.0K    [redacted]roperties
4.0K    [redacted]ry-point-offset-checkpoint
4.0K    [redacted]ation-offset-checkpoint
4.0K    [redacted]-3
4.0K    [redacted]-4
4.0K    [redacted]-5
4.0K    [redacted]-7
4.0K    [redacted]-9
[root@node-str-coreetHK kafka-logs]# pwd
/srv/BigData/streaming/data3/kafka-logs
```

Step 6 Determine the method of changing the data retention period. The default global data retention period of Kafka is seven days. A large amount of data may be written to some topics, and these topics reside on the partitions on the disk with high usage.

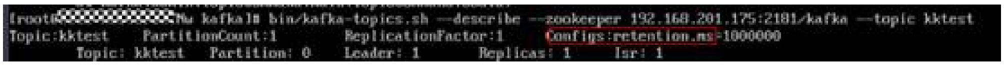
- You can change the global data retention period to a smaller value to release disk space. This method requires a Kafka service restart, which may affect service running. For details, see **Step 7**.
- You can change the data retention period of a single topic to a smaller value to release disk space. This configuration takes effect without a Kafka service restart. For details, see **Step 8**.

Step 7 Log in to Manager. On the Kafka service configuration page, switch to **All Configurations** and search for the **log.retention.hours** configuration item. The default value is 7 days. Change it based on the site requirements.

Step 8 Change the data retention time of the topics on these disks.

1. Check the retention time of the topic data.

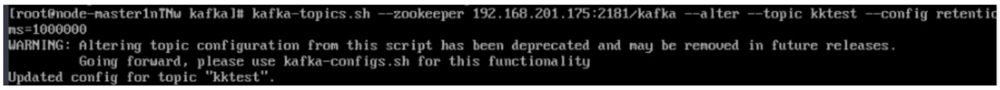
   bin/kafka-topics.sh --describe --zookeeper ***<ZooKeeper cluster service IP address>*****:2181/kafka --topic kktest**

   ```
   [root@[redacted]Nw kafka]# bin/kafka-topics.sh --describe --zookeeper 192.168.201.175:2181/kafka --topic kktest
   Topic:kktest    PartitionCount:1        ReplicationFactor:1     Configs:retention.ms=1000000
           Topic: kktest   Partition: 0    Leader: 1       Replicas: 1     Isr: 1
   ```

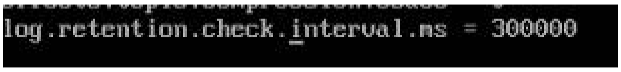
2. Set the topic data retention time. **--topic** indicates the topic name, and **retention.ms** indicates the data retention time, in milliseconds.

   kafka-topics.sh --zookeeper ***<ZooKeeper cluster service IP address>*****:2181/kafka --alter --topic kktest --config retention.ms=1000000**

   ```
   [root@node-master1nTNw kafka]# kafka-topics.sh --zookeeper 192.168.201.175:2181/kafka --alter --topic kktest --config retentic
   ms=1000000
   WARNING: Altering topic configuration from this script has been deprecated and may be removed in future releases.
            Going forward, please use kafka-configs.sh for this functionality
   Updated config for topic "kktest".
   ```

   After the data retention time is set, the deletion operation may not be performed immediately. The deletion operation starts after the time specified by **log.retention.check.interval.ms**. You can check whether the **delete** field exists in the **server.log** file of Kafka to determine whether the deletion operation takes effect. If the **delete** field exists, the deletion operation has taken effect. You can also run the **df –h** command to check the disk usage and determine whether the setting takes effect.

   ```
   log.retention.check.interval.ms = 300000
   ```

----End

15.14 Using Oozie

15.14.1 Oozie Jobs Do Not Run When a Large Number of Jobs Are Submitted Concurrently

Issue

When a large number of Oozie jobs are submitted concurrently, the jobs do not run.

Symptom

When a large number of Oozie jobs are submitted concurrently, the jobs do not run.

Cause Analysis

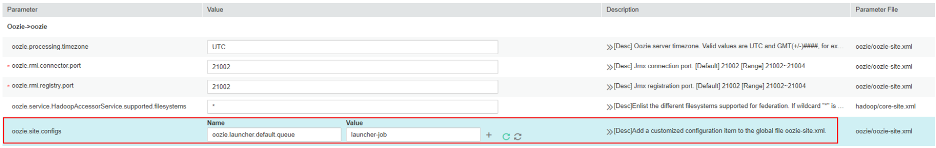
When Oozie submits a job, an oozie-launcher job is started first, and then the oozie-launcher job submits the real job for execution. By default, the oozie-launcher job and the real job are in the same queue.

When a large number of Oozie jobs are submitted concurrently, a large number of oozie-launcher jobs may be started, exhausting the resources of the queue. As a result, no more resources are available to start real jobs, and the jobs are not executed.

Procedure

Step 1 Create a queue for Oozie. For details, see **User Guide** > **Managing an Existing Cluster** > **Tenant Management** > **Creating a Tenant**. You can also use the launcher-job queue generated during MRS cluster creation.

Step 2 On Manager, choose **Cluster** > **Services** > **Oozie** > **Configurations**, search for **oozie.site.configs**, and add **oozie.launcher.default.queue** as the parameter name and **launcher-job** as the value.



Parameter	Value	Description	Parameter File
Oozie->oozie			
oozie.processing.timezone	UTC	»[Desc] Oozie server timezone. Valid values are UTC and GMT(+/-)####, for ex...	oozie/oozie-site.xml
* oozie.rmi.connector.port	21002	»[Desc] Jmx connection port. [Default] 21002 [Range] 21002~21004	oozie/oozie-site.xml
* oozie.rmi.registry.port	21002	»[Desc] Jmx registration port. [Default] 21002 [Range] 21002~21004	oozie/oozie-site.xml
oozie.service.HadoopAccessorService.supported.filesystems	*	»[Desc]Enlist the different filesystems supported for federation. If wildcard "*" is ...	hadoop/core-site.xml
oozie.site.configs	Name: oozie.launcher.default.queue Value: launcher-job	»[Desc]Add a customized configuration item to the global file oozie-site.xml.	oozie/oozie-site.xml

----End

15.15 Using Presto

15.15.1 During sql-standard-with-group Configuration, a Schema Fails to Be Created and the Error Message "Access Denied" Is Displayed

Issue

A schema fails to be created during sql-standard-with-group configuration and the error message "Access Denied" is displayed.

Symptom

```
CREATE SCHEMA hive.sf2 WITH (location = 'obs://obs-zy1234/sf2');Query 20200224_031203_00002_g6gzy
failed: Access Denied: Cannot create schema sf2
```

Cause Analysis

To create a schema in Presto, you must have the administrator permission of Hive.

Procedure

MRS Manager:

- Method 1:
  a. Log in to MRS Manager and choose **System** > **Manage User**.
  b. Locate the row that contains the target user, and click **Modify** in the **Operation** column.
  c. Click **Select and Add Role** to assign the **System_administrator** permission to the user.
  d. Click **OK**.
- Method 2:
  a. Log in to MRS Manager and choose **System** > **Manage Role**.
  b. Click **Create Role** and set the following parameters:
     - Enter a role name, for example, **hive_admin**.
     - Set **Permission** to **Hive** and select **Hive Admin Privilege**.
  c. Click **OK** to save the role.
  d. Choose **System** > **Manage User**.
  e. Locate the row that contains the target user, and click **Modify** in the **Operation** column.
  f. Click **Select and Add Role** to add the newly created hive_admin permission to the user.
  g. Click **OK**.

FusionInsight Manager:

- Method 1:
  a. Log in to FusionInsight Manager and choose **System** > **Permission** > **User**.
  b. Locate the row that contains the target user, and click **Modify** in the **Operation** column.
  c. Click **Add** next to the role to assign the **System_administrator** permission to the user.
  d. Click **OK**.
- Method 2:
  a. Log in to FusionInsight Manager and choose **System** > **Permission** > **Role**.
  b. Click **Create Role** and set the following parameters:
     - Enter a role name, for example, **hive_admin**.
     - To configure resource permissions, select **Hive** and **Hive Admin Permissions**.
  c. Click **OK** to save the role.
  d. Choose **System** > **Permission** > **User**.

e. Locate the row that contains the target user, and click **Modify** in the **Operation** column.

f. Click **Add** next to the role to add the **hive_admin** permission for the user.

g. Click **OK**.

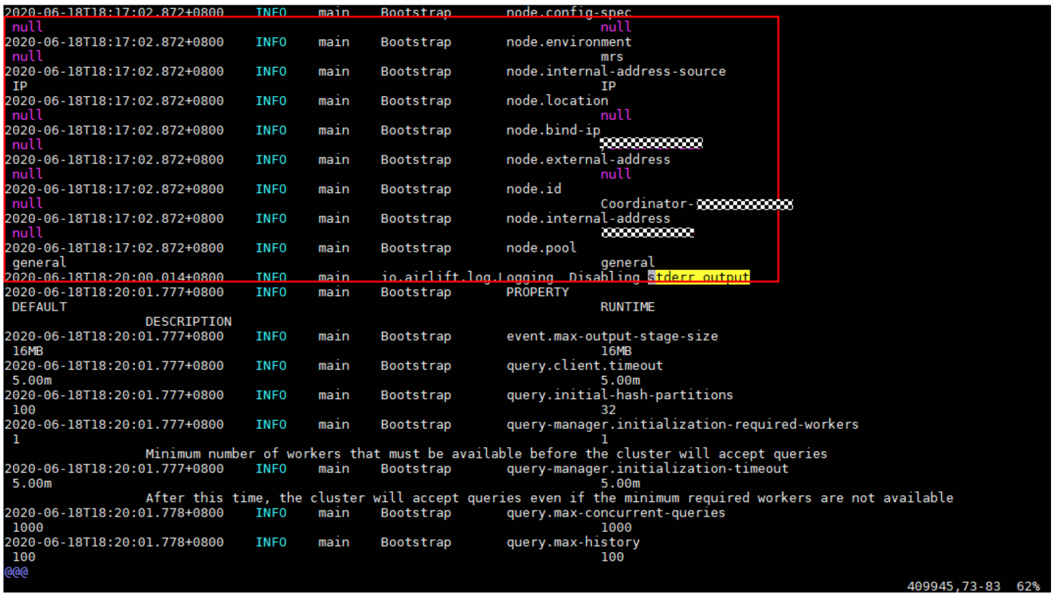
15.15.2 The Presto coordinator cannot be started properly.

Issue

The coordinator process of Presto is killed due to an unknown reason, or the coordinator process of Presto cannot be started.

Symptom

The Presto coordinator process cannot be started properly. On the Manager page, it is shown that the presto coordinator process is started properly and its status is normal. However, the background log shows that the coordinator process is not started. Only the following log is displayed:



```
2020-06-18T18:17:02.872+0800    INFO    main    Bootstrap       node.config-spec
 null                                                            null
2020-06-18T18:17:02.872+0800    INFO    main    Bootstrap       node.environment
 null                                                            mrs
2020-06-18T18:17:02.872+0800    INFO    main    Bootstrap       node.internal-address-source
 IP                                                              IP
2020-06-18T18:17:02.872+0800    INFO    main    Bootstrap       node.location
 null                                                            null
2020-06-18T18:17:02.872+0800    INFO    main    Bootstrap       node.bind-ip
 null                                                            [redacted]
2020-06-18T18:17:02.872+0800    INFO    main    Bootstrap       node.external-address
 null                                                            null
2020-06-18T18:17:02.872+0800    INFO    main    Bootstrap       node.id
 null                                                            Coordinator-[redacted]
2020-06-18T18:17:02.872+0800    INFO    main    Bootstrap       node.internal-address
 null                                                            [redacted]
2020-06-18T18:17:02.872+0800    INFO    main    Bootstrap       node.pool
 general                                                         general
2020-06-18T18:20:00.014+0800    INFO    main    io.airlift.log.Logging  Disabling stderr output
2020-06-18T18:20:01.777+0800    INFO    main    Bootstrap       PROPERTY
 DEFAULT                                                         RUNTIME
                 DESCRIPTION
2020-06-18T18:20:01.777+0800    INFO    main    Bootstrap       event.max-output-stage-size
 16MB                                                            16MB
2020-06-18T18:20:01.777+0800    INFO    main    Bootstrap       query.client.timeout
 5.00m                                                           5.00m
2020-06-18T18:20:01.777+0800    INFO    main    Bootstrap       query.initial-hash-partitions
 100                                                             32
2020-06-18T18:20:01.777+0800    INFO    main    Bootstrap       query-manager.initialization-required-workers
 1                                                               1
                 Minimum number of workers that must be available before the cluster will accept queries
2020-06-18T18:20:01.777+0800    INFO    main    Bootstrap       query-manager.initialization-timeout
 5.00m                                                           5.00m
                 After this time, the cluster will accept queries even if the minimum required workers are not available
2020-06-18T18:20:01.778+0800    INFO    main    Bootstrap       query.max-concurrent-queries
 1000                                                            1000
2020-06-18T18:20:01.778+0800    INFO    main    Bootstrap       query.max-history
 100                                                             100
@@@
                                                                                              409945,73-83   62%
```

The Presto coordinator is killed before being started, and no other logs are printed. Other Presto logs do not indicate the reason why the presto coordinator is killed.

Cause Analysis

The port check logic of the presto health check script does not distinguish ports.

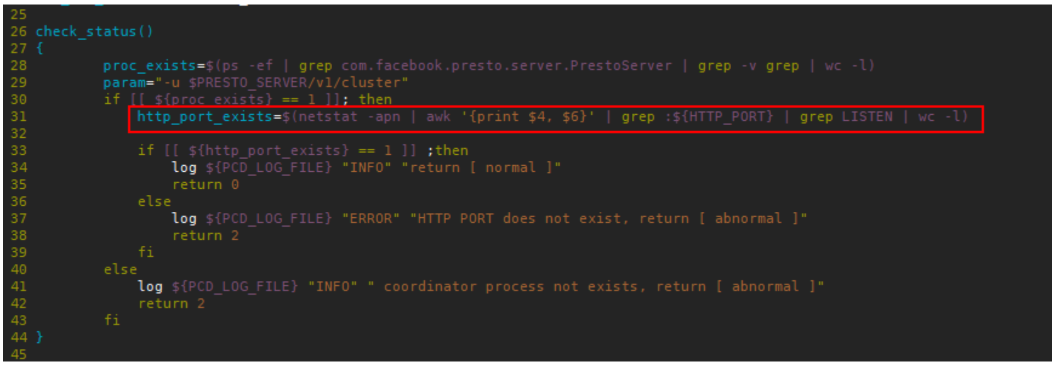
Procedure

Step 1 Use a tool to log in to the master nodes of the cluster and perform the following operations:

Step 2 Run the following command to edit the file:

vim /opt/Bigdata/MRS_xxx/install/FusionInsight-Presto-*/ha/module/harm/plugin/script/pcd.sh

Change line 31 in the file to **http_port_exists=$(netstat -apn | awk '{print $4, $6}' | grep :${HTTP_PORT} | grep LISTEN | wc -l)**.



```
25
26 check_status()
27 {
28         proc_exists=$(ps -ef | grep com.facebook.presto.server.PrestoServer | grep -v grep | wc -l)
29         param="-u $PRESTO_SERVER/v1/cluster"
30         if [[ ${proc_exists} == 1 ]]; then
31             http_port_exists=$(netstat -apn | awk '{print $4, $6}' | grep :${HTTP_PORT} | grep LISTEN | wc -l)
32
33             if [[ ${http_port_exists} == 1 ]] ;then
34                 log ${PCD_LOG_FILE} "INFO" "return [ normal ]"
35                 return 0
36             else
37                 log ${PCD_LOG_FILE} "ERROR" "HTTP PORT does not exist, return [ abnormal ]"
38                 return 2
39             fi
40         else
41             log ${PCD_LOG_FILE} "INFO" " coordinator process not exists, return [ abnormal ]"
42             return 2
43         fi
44 }
45
```

Step 3 Save the modification. On FusionInsight Manager, choose **Services** > **Presto** > **Instances** to restart the Coordinator process.

----End

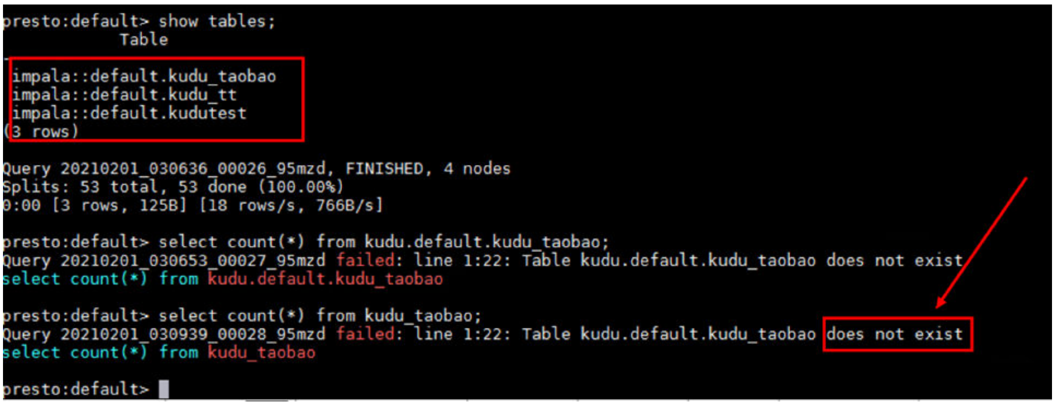
15.15.3 An Error Is Reported When Presto Is Used to Query a Kudu Table

Issue

An error is reported when Presto is used to query a Kudu table.

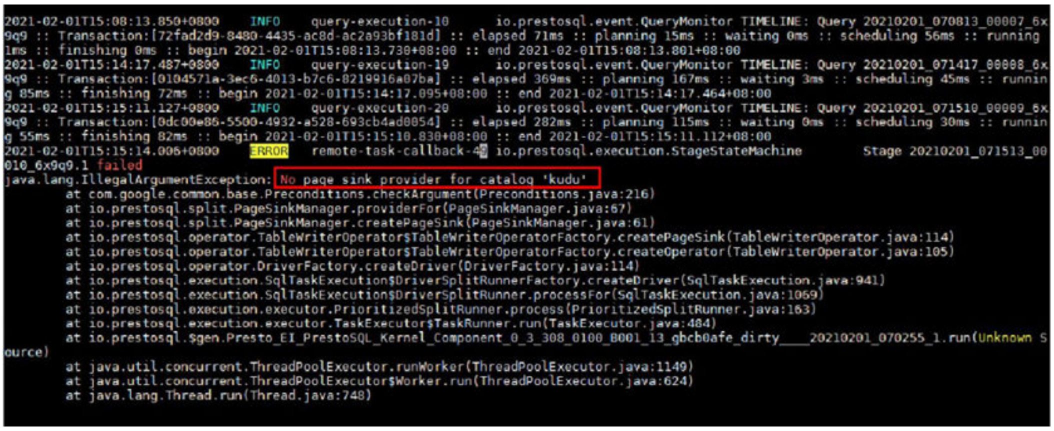
Symptom

When Presto is used to query a Kudu table, the following error message is displayed.



```
presto:default> show tables;
              Table
-
 impala::default.kudu_taobao
 impala::default.kudu_tt
 impala::default.kudutest
(3 rows)

Query 20210201_030636_00026_95mzd, FINISHED, 4 nodes
Splits: 53 total, 53 done (100.00%)
0:00 [3 rows, 125B] [18 rows/s, 766B/s]

presto:default> select count(*) from kudu.default.kudu_taobao;
Query 20210201_030653_00027_95mzd failed: line 1:22: Table kudu.default.kudu_taobao does not exist
select count(*) from kudu.default.kudu_taobao

presto:default> select count(*) from kudu_taobao;
Query 20210201_030939_00028_95mzd failed: line 1:22: Table kudu.default.kudu_taobao does not exist
select count(*) from kudu_taobao

presto:default>
```

Error information



```
2021-02-01T15:08:13.850+0800    INFO    query-execution-10      io.prestosql.event.QueryMonitor TIMELINE: Query 20210201_070813_00007_6x
9q9 :: Transaction:[72fad2d9-8480-4435-ac8d-ac2a93bf181d] :: elapsed 71ms :: planning 15ms :: waiting 0ms :: scheduling 56ms :: running
1ms :: finishing 0ms :: begin 2021-02-01T15:08:13.730+08:00 :: end 2021-02-01T15:08:13.801+08:00
2021-02-01T15:14:17.487+0800    INFO    query-execution-19      io.prestosql.event.QueryMonitor TIMELINE: Query 20210201_071417_00008_6x
9q9 :: Transaction:[0104571a-3ec6-4013-b7c6-8219916a07ba] :: elapsed 369ms :: planning 167ms :: waiting 3ms :: scheduling 45ms :: runnin
g 85ms :: finishing 72ms :: begin 2021-02-01T15:14:17.095+08:00 :: end 2021-02-01T15:14:17.464+08:00
2021-02-01T15:15:11.127+0800    INFO    query-execution-20      io.prestosql.event.QueryMonitor TIMELINE: Query 20210201_071510_00009_6x
9q9 :: Transaction:[0dc00e86-5500-4932-a528-693cb4ad0054] :: elapsed 282ms :: planning 115ms :: waiting 0ms :: scheduling 30ms :: runnin
g 55ms :: finishing 82ms :: begin 2021-02-01T15:15:10.830+08:00 :: end 2021-02-01T15:15:11.112+08:00
2021-02-01T15:15:14.006+0800    ERROR   remote-task-callback-4  io.prestosql.execution.StageStateMachine        Stage 20210201_071513_00
010_6x9q9.1 failed
java.lang.IllegalArgumentException: No page sink provider for catalog 'kudu'
        at com.google.common.base.Preconditions.checkArgument(Preconditions.java:216)
        at io.prestosql.split.PageSinkManager.providerFor(PageSinkManager.java:67)
        at io.prestosql.split.PageSinkManager.createPageSink(PageSinkManager.java:61)
        at io.prestosql.operator.TableWriterOperator$TableWriterOperatorFactory.createPageSink(TableWriterOperator.java:114)
        at io.prestosql.operator.TableWriterOperator$TableWriterOperatorFactory.createOperator(TableWriterOperator.java:105)
        at io.prestosql.operator.DriverFactory.createDriver(DriverFactory.java:114)
        at io.prestosql.execution.SqlTaskExecution$DriverSplitRunnerFactory.createDriver(SqlTaskExecution.java:941)
        at io.prestosql.execution.SqlTaskExecution$DriverSplitRunner.processFor(SqlTaskExecution.java:1069)
        at io.prestosql.execution.executor.PrioritizedSplitRunner.process(PrioritizedSplitRunner.java:163)
        at io.prestosql.execution.executor.TaskExecutor$TaskRunner.run(TaskExecutor.java:484)
        at io.prestosql.$gen.Presto_EI_PrestoSQL_Kernel_Component_0_3_308_0100_B001_13_gbcb0afe_dirty____20210201_070255_1.run(Unknown S
ource)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
        at java.lang.Thread.run(Thread.java:748)
```

Cause Analysis

There are no Kudu configurations on the actually running node (node where the worker instance is located).

Procedure

Step 1 Add configuration file **kudu.properties** to all worker instance nodes in the Presto cluster.

Path for storing the configuration file: **/opt/Bigdata/MRS_xxx/1_x_Worker/etc/catalog/** (Change the path based on the actual cluster version.)

Configuration file content:

```
connector.name=kudu
kudu.client.master-addresses=KuduMasterIP1:port,KuduMasterIP2:port,KuduMasterIP3:port
```

NOTE

- Set the IP address and port number of the KuduMaster node based on the site requirements.
- Add the file permission and owner group that are the same as those of other files in the file save path to the configuration file.

Step 2 After the modification, choose **Components** > **Kudu** on the cluster details page, click **More**, and select **Restart Service**.

----End

15.15.4 No Data is Found in the Hive Table Using Presto

Issue

When Presto is used to query the Hive table, no data is found.

Symptom

Presto cannot query the data written by **union** statements executed by the Tez engine.

Cause Analysis

When Hive uses the Tez engine to execute the **union** statements, the output file is stored in the **HIVE_UNION_SUBDIR** directory. However, Presto does not access files in child directories by default. Therefore, data in the **HIVE_UNION_SUBDIR** directory is not read.

Procedure

Step 1 On the MRS console, click the cluster name, and choose **Components** > **Presto** > **Service Configuration**.

Step 2 Change **Basic** to **All**.

Step 3 In the navigation tree on the left, choose **Presto** > **Hive**. In the **catalog/hive.properties** file, add the **hive.recursive-directories** parameter and set it to **true**.

Step 4 Click **Save Configuration** and select **Restart the affected services or instances**.

----End

15.16 Using Spark

15.16.1 An Error Occurs When the Split Size Is Changed in a Spark Application

Issue

An error occurs when the split size is changed in a Spark application.

Symptom

A user needs to implement multiple mappers by changing the maximum split size to make the Spark application run faster. However, an error occurs when the user runs the **set $Parameter** command to modify the Hive configuration.

```
0: jdbc:hive2://192.168.1.18:21066/> set mapred.max.split.size=1000000;
Error: Error while processing statement: Cannot nodify mapred.max.split.size at runtime. It is not in list of
params that are allowed to be modified at runtime( state=42000,code=1)
```

Cause Analysis

- Before the **hive.security.whitelist.switch** parameter is set to enable or disable the whitelist in security mode, the allowed parameters must have been configured in **hive.security.authorization.sqlstd.confwhitelist**.
- The default whitelist does not contain the **mapred.max.split.size** parameter. Therefore, the system displays a message indicating that the maximum split size cannot be changed.

Procedure

Step 1 Go to the Hive configuration page.

- For versions earlier than MRS 2.0.1: Log in to MRS Manager, choose **Services** > **Hive** > **Service Configuration**, and select **All** from the **Basic** drop-down list.
- For MRS 2.0.1 or later: Click the cluster name on the MRS console, choose **Components** > **Hive** > **Service Configuration**, and select **All** from the **Basic** drop-down list.

  NOTE

  If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)
- For MRS 3.*x* or later: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the target cluster and choose **Services** > **Hive** > **Configurations** > **All Configurations**.

Step 2 Search for **hive.security.authorization.sqlstd.confwhitelist.append**, and add **mapred.max.split.size** to **hive.security.authorization.sqlstd.confwhitelist.append**. For details, see **Component Operation Guide** > **Using Hive** > **Using Hive from Scratch**.

Step 3 Save the configuration and restart the Hive component.

Step 4 Run the **set mapred.max.split.size=1000000;** command. If no error occurs, the modification is successful.

----End

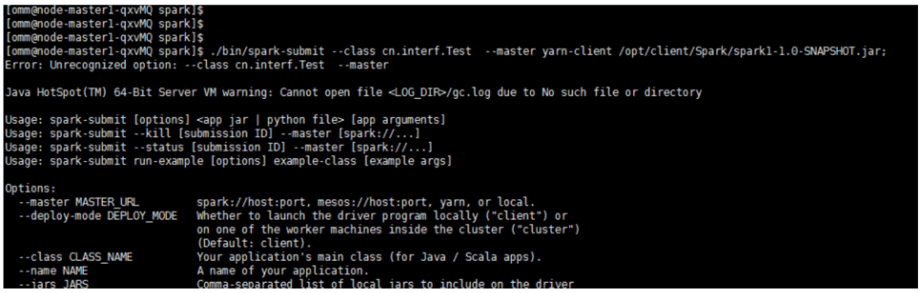
15.16.2 An Error Is Reported When Spark Is Used

Issue

When Spark is used, the cluster fails to run.

Symptom

When Spark is used, the cluster fails to run.



```
[omm@node-master1-qxvMQ spark]$
[omm@node-master1-qxvMQ spark]$
[omm@node-master1-qxvMQ spark]$
[omm@node-master1-qxvMQ spark]$ ./bin/spark-submit --class cn.interf.Test  --master yarn-client /opt/client/Spark/spark1-1.0-SNAPSHOT.jar;
Error: Unrecognized option: --class cn.interf.Test  --master

Java HotSpot(TM) 64-Bit Server VM warning: Cannot open file <LOG_DIR>/gc.log due to No such file or directory

Usage: spark-submit [options] <app jar | python file> [app arguments]
Usage: spark-submit --kill [submission ID] --master [spark://...]
Usage: spark-submit --status [submission ID] --master [spark://...]
Usage: spark-submit run-example [options] example-class [example args]

Options:
  --master MASTER_URL         spark://host:port, mesos://host:port, yarn, or local.
  --deploy-mode DEPLOY_MODE   Whether to launch the driver program locally ("client") or
                              on one of the worker machines inside the cluster ("cluster")
                              (Default: client).
  --class CLASS_NAME          Your application's main class (for Java / Scala apps).
  --name NAME                 A name of your application.
  --jars JARS                 Comma-separated list of local jars to include on the driver
```

Cause Analysis

- Invalid characters are added during command execution.
- The owner and owner group of the uploaded JAR file is incorrect.

Procedure

Step 1 Run **./bin/spark-submit --class cn.interf.Test --master yarn-client /opt/client/Spark/spark1-1.0-SNAPSHOT.jar;** to check whether invalid characters are imported.

Step 2 If they are imported, modify the invalid characters and run the command again.

Step 3 After the command is executed again, other errors occur. Both the owner and the owner group of the JAR file are **root**.

Step 4 Change the owner and the owner group of the JAR file to **omm:wheel**.

----End

15.16.3 A Spark Job Fails to Run Due to Incorrect JAR File Import

Issue

A Spark job fails to be executed.

Symptom

A Spark job fails to be executed.

Cause Analysis

The imported JAR file is incorrect when the Spark job is executed. As a result, the Spark job fails to be executed.

Procedure

Step 1 Log in to any Master node.

Step 2 Run the **cd /opt/Bigdata/MRS_*/install/FusionInsight-Spark-*/spark/examples/jars** command to view the JAR file of the sample program.

NOTE

A JAR file name contains a maximum of 1023 characters and cannot include special characters (;|&>,<'$). In addition, it cannot be left blank or full of spaces.

Step 3 Check the executable programs in the OBS bucket. The executable programs can be stored in HDFS or OBS. The paths vary according to file systems.

NOTE

- OBS storage path: starts with **obs://**, for example, **obs://wordcount/program/hadoop-mapreduce-examples-2.7.x.jar**.
- HDFS storage path: starts with **/user**. Spark Script must end with **.sql**, and MR and Spark must end with **.jar**. The **.sql** and **.jar** are case-insensitive.

----End

15.16.4 A Spark Job Is Pending Due to Insufficient Memory

Issue

Memory is insufficient to submit a Spark job. As a result, the job is in the pending state for a long time or out of memory (OMM) occurs during job running.

Symptom

The job is pending for a long time after being submitted. The following error information is displayed after the job is executed repeatedly:

```
Exception in thread "main" org.apache.spark.SparkException: Job aborted due to stage failure:
Aborting TaskSet 3.0 because task 0 (partition 0) cannot run anywhere due to node and executor blacklist.
Blacklisting behavior can be configured via spark.blacklist.*.
```

Cause Analysis

The memory is insufficient. As a result, the submitted Spark job is in the pending state for a long time.

Procedure

Step 1 Log in to the MRS console, click a cluster name on the **Active Clusters** page and view the node specifications of the cluster on the **Nodes** tab page.

Step 2 Add cluster resources owned by the **nodemanager** process.

MRS Manager:

1. Log in to MRS Manager and choose **Services** > **Yarn** > **Service Configuration**.
2. Set **Type** to **All**, and then search for **yarn.nodemanager.resource.memory-mb** in the search box to view the value of this parameter. You are advised to set the parameter value to 75% to 90% of the total physical memory of nodes.

FusionInsight Manager:

1. Log in to FusionInsight Manager. Choose **Cluster** > **Service** > **Yarn**.
2. Choose **Configurations** > **All Configurations**. Search for **yarn.nodemanager.resource.memory-mb** in the search box and check the parameter value. You are advised to set the parameter value to 75% to 90% of the total physical memory of nodes.

Step 3 Modify the Spark service configuration.

MRS Manager:

1. Log in to MRS Manager and choose **Services** > **Spark** > **Service Configuration**.
2. Set **Type** to **All**, and then search for **spark.driver.memory** and **spark.executor.memory** in the search box.

   Set these parameters to a larger or smaller value based on the complexity and memory requirements of the submitted Spark job. (Generally, the values need to be increased.)

FusionInsight Manager:

1. Log in to FusionInsight Manager. Choose **Cluster** > **Service** > **Spark**.
2. Choose **Configurations** > **All Configurations**. Search for **spark.driver.memory** and **spark.executor.memory** in the search box and increase or decrease the values based on actual requirements. Generally, increase the values based on the complexity and memory of the submitted Spark job.

NOTE

- If a SparkJDBC job is used, search for **SPARK_EXECUTOR_MEMORY** and **SPARK_DRIVER_MEMORY** and modify their values based on the complexity and memory requirements of the submitted Spark job. (Generally, the values need to be increased.)
- If the number of cores needs to be specified, you can search for **spark.driver.cores** and **spark.executor.cores** and modify their values.

Step 4 Scale out the cluster if the preceding requirements still cannot be met because Spark depends on the memory for computing.

----End

15.16.5 An Error Is Reported During Spark Running

Issue

The specified class cannot be found when a Spark job is running.

Symptom

The specified class cannot be found when a Spark job is running. The error message is as follows:

```
Exception encountered | org.apache.spark.internal.Logging$class.logError(Logging.scala:91)
org.apache.hadoop.hbase.DoNotRetryIOException: java.lang.ClassNotFoundException:
org.apache.phoenix.filter.SingleCQKeyValueComparisonFilter
```

Cause Analysis

The default path configured by the user is incorrect.

Procedure

Step 1 Log in to any Master node.

Step 2 Modify the configuration file in the Spark client directory.

Run the **vim /opt/client/Spark/spark/conf/spark-defaults.conf** command to open the **spark-defaults.conf** file and set **spark.executor.extraClassPath** to **${PWD}/***.

----End

15.16.6 Executor Memory Reaches the Threshold Is Displayed in Driver

Symptom

A Spark task fails to be submitted due to excessive memory usage.

Cause Analysis

The Driver log prints that the applied Executor memory exceeds the cluster limit.

```
16/02/06 14:11:25 INFO Client: Verifying our application has not requested more than the maximum
memory capability of the cluster (6144 MB per container)
16/02/06 14:11:29 ERROR SparkContext: Error initializing SparkContext.
java.lang.IllegalArgumentException: Required executor memory (10240+1024 MB) is above the max
threshold (6144 MB) of this cluster!
```

Spark tasks are submitted to Yarn and the resources used by the Executor to run tasks are managed by Yarn. From the error message, you can see that when a user starts the Executor, 10 GB memory is specified, which exceeds the upper memory limit of each Container set by Yarn. As a result, the task cannot be started.

Solution

Modify the Yarn configuration to increase the restriction on containers. For example, you can adjust parameter **yarn.scheduler.maximum-allocation-mb** to control the resources for starting the Executor. Restart the Yarn service after the modification.

You can modify the configuration as follows:

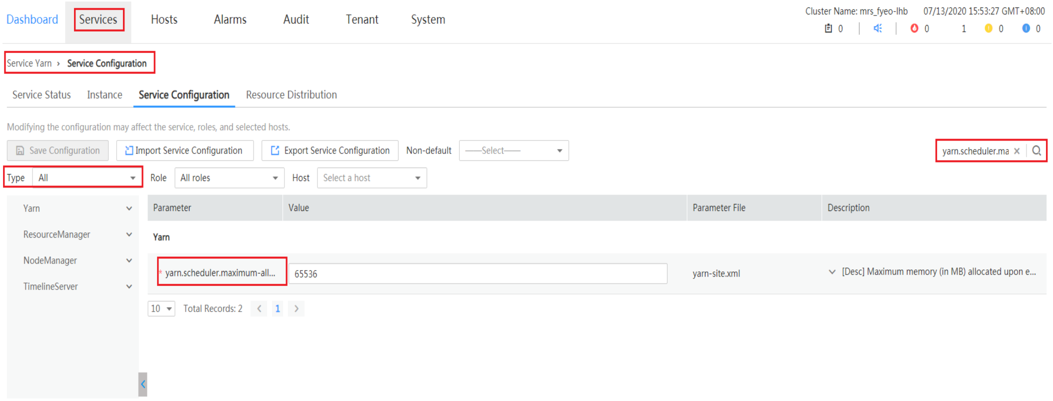
MRS Manager:

Step 1 Log in to MRS Manager.

Step 2 Choose **Services** > **Yarn** > **Service Configuration** and set **Type** to **All**.

Step 3 In **Search**, enter **yarn.scheduler.maximum-allocation-mb** to modify the parameter, save the configuration, and then restart the service. See the following figure.

Figure 15-56 Modifying Yarn service parameters



Cluster Name: mrs_fyeo-lhb 07/13/2020 15:53:27 GMT+08:00

Modifying the configuration may affect the service, roles, and selected hosts.

Non-default ——Select——

yarn.scheduler.ma

Type All Role All roles Host Select a host

Parameter	Value	Parameter File	Description
Yarn			
* yarn.scheduler.maximum-all...	65536	yarn-site.xml	[Desc] Maximum memory (in MB) allocated upon e...

Total Records: 2

----End

FusionInsight Manager:

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > **Service** > **Yarn**. Click **Configurations** and select **All Configurations**.

Step 3 In **Search**, enter **yarn.scheduler.maximum-allocation-mb** to modify the parameter, save the configuration, and then restart the service.

----End

15.16.7 Message "Can't get the Kerberos realm" Is Displayed in Yarn-cluster Mode

Symptom

A Spark task fails to be submitted due to an authentication failure.

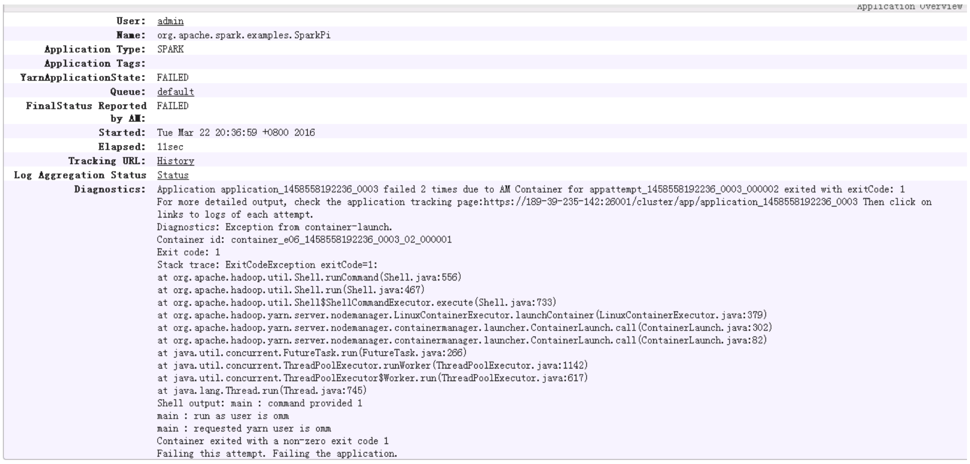
Cause Analysis

1. According to the exception printed in the driver log, the token used to connect to HDFS cannot be found.

```
16/03/22 20:37:10 WARN Client: Exception encountered while connecting to the server :
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretManager
$InvalidToken): token (HDFS_DELEGATION_TOKEN token 192 for admin) can't be found in cache
16/03/22 20:37:10 WARN Client: Failed to cleanup staging dir .sparkStaging/
application_1458558192236_0003
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretManager
$InvalidToken): token (HDFS_DELEGATION_TOKEN token 192 for admin) can't be found in cache
```

2. The native Yarn web UI shows that ApplicationMaster fails to be started twice and the task exits.

Figure 15-57 ApplicationMaster start failure



```
                 User:  admin
                 Name:  org.apache.spark.examples.SparkPi
     Application Type:  SPARK
     Application Tags:
YarnApplicationState:  FAILED
                Queue:  default
FinalStatus Reported    FAILED
               by AM:
              Started:  Tue Mar 22 20:36:59 +0800 2016
              Elapsed:  11sec
         Tracking URL:  History
Log Aggregation Status  Status
          Diagnostics:  Application application_1458558192236_0003 failed 2 times due to AM Container for appattempt_1458558192236_0003_000002 exited with exitCode: 1
                        For more detailed output, check the application tracking page:https://189-39-235-142:26001/cluster/app/application_1458558192236_0003 Then click on
                        links to logs of each attempt.
                        Diagnostics: Exception from container-launch.
                        Container id: container_e06_1458558192236_0003_02_000001
                        Exit code: 1
                        Stack trace: ExitCodeException exitCode=1:
                        at org.apache.hadoop.util.Shell.runCommand(Shell.java:556)
                        at org.apache.hadoop.util.Shell.run(Shell.java:467)
                        at org.apache.hadoop.util.Shell$ShellCommandExecutor.execute(Shell.java:733)
                        at org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor.launchContainer(LinuxContainerExecutor.java:379)
                        at org.apache.hadoop.yarn.server.nodemanager.containermanager.launcher.ContainerLaunch.call(ContainerLaunch.java:302)
                        at org.apache.hadoop.yarn.server.nodemanager.containermanager.launcher.ContainerLaunch.call(ContainerLaunch.java:82)
                        at java.util.concurrent.FutureTask.run(FutureTask.java:266)
                        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1142)
                        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:617)
                        at java.lang.Thread.run(Thread.java:745)
                        Shell output: main : command provided 1
                        main : run as user is omm
                        main : requested yarn user is omm
                        Container exited with a non-zero exit code 1
                        Failing this attempt. Failing the application.
```

3. The ApplicationMaster log shows the following error information:

```
Exception in thread "main" java.lang.ExceptionInInitializerError
Caused by: org.apache.spark.SparkException: Unable to load YARN support
Caused by: java.lang.IllegalArgumentException: Can't get Kerberos realm
Caused by: java.lang.reflect.InvocationTargetException
Caused by: KrbException: Cannot locate default realm
Caused by: KrbException: Generic error (description in e-text) (60) - Unable to locate Kerberos realm
org.apache.hadoop.hive.metastore.MetaStoreUtils.newInstance(MetaStoreUtils.java:1410)
```

```
... 86 more
Caused by: javax.jdo.JDOFatalInternalException: Unexpected exception caught.
NestedThrowables:java.lang.reflect.InvocationTargetException
... 110 more
```

4. When you execute **./spark-submit --class yourclassname --master yarn-cluster /yourdependencyjars** to submit a task in Yarn-cluster mode, the driver is enabled in the cluster. Because the client's **spark.driver.extraJavaOptions** is loaded, you cannot find the **kdc.conf** file in the target path on the cluster node and cannot obtain information required for Kerberos authentication. As a result, the ApplicationMaster fails to be started.

Solution

When submitting a task on the client, configure the **spark.driver.extraJavaOptions** parameter in the CLI. In this way, the **spark.driver.extraJavaOptions** parameter in the **spark-defaults.conf** file is not automatically loaded from the client path. When starting a Spark task, use **--conf** to specify the driver configuration as follows (note that the quotation mark after **spark.driver.extraJavaOptions=** is mandatory):

./spark-submit -class yourclassname --master yarn-cluster --conf spark.driver.extraJavaOptions="

-Dlog4j.configuration=file:/opt/client/Spark/spark/conf/log4j.properties -Djetty.version=x.y.z -Dzookeeper.server.principal=zookeeper/hadoop.794bbab6_9505_44cc_8515_b4eddc84e6c1.com -Djava.security.krb5.conf=/opt/client/KrbClient/kerberos/var/krb5kdc/krb5.conf -Djava.security.auth.login.config=/opt/client/Spark/spark/conf/jaas.conf -Dorg.xerial.snappy.tempdir=/opt/client/Spark/tmp -Dcarbon.properties.filepath=/opt/client/Spark/spark/conf/carbon.properties" ../yourdependencyjars

15.16.8 Failed to Start spark-sql and spark-shell Due to JDK Version Mismatch

Symptom

The JDK version does not match. As a result, the client fails to start spark-sql and spark-shell.

Cause Analysis

1. The following error information is displayed on the Driver:

```
 Exception Occurs: BadPadding 16/02/22 14:25:38 ERROR Schema: Failed initialising database. Unable
to open a test connection to the given database. JDBC url = jdbc:postgresql://ip:port/sparkhivemeta,
username = spark. Terminating connection pool (set lazyInit to true if you expect to start your
database after your app).
```

2. When a SparkSQL task is used, DBService needs to be accessed to obtain metadata information. On the client, the ciphertext needs to be decrypted for access. During the use, the user does not follow the process or configure environment variables, and the default JDK version exists in the environment variables of the client. As a result, the decryption program invoked during decryption is abnormal, and the user is locked.

Solution

Step 1 Run the **which java** command to check whether the default Java command is the Java command of the client.

Step 2 If it is not, go to the next step.

source ${client_path}/bigdata_env

Run the **kinit** *username* command and enter the password corresponding to the username to start the task.

----End

15.16.9 ApplicationMaster Failed to Start Twice in Yarn-client Mode

Symptom

In Yarn-client mode, ApplicationMaster fails to start twice.

Cause Analysis

1. Driver exception:

```
16/05/11 18:10:56 INFO Client:
client token: N/A
diagnostics: Application application_1462441251516_0024 failed 2 times due to AM Container for
appattempt_1462441251516_0024_000002 exited with  exitCode: 10
For more detailed output, check the application tracking page:https://hdnode5:26001/cluster/app/
application_1462441251516_0024 Then click on links to logs of each attempt.
Diagnostics: Exception from container-launch.
Container id: container_1462441251516_0024_02_000001
```

2. The ApplicationMaster log file contains the following error information:

```
2016-05-12 10:21:23,715 | ERROR | [main] | Failed to connect to driver at 192.168.30.57:23867,
retrying ... | org.apache.spark.Logging$class.logError(Logging.scala:75)
2016-05-12 10:21:24,817 | ERROR | [main] | Failed to connect to driver at 192.168.30.57:23867,
retrying ... | org.apache.spark.Logging$class.logError(Logging.scala:75)
2016-05-12 10:21:24,918 | ERROR | [main] | Uncaught exception:  | org.apache.spark.Logging
$class.logError(Logging.scala:96)
org.apache.spark.SparkException: Failed to connect to driver!
at org.apache.spark.deploy.yarn.ApplicationMaster.waitForSparkDriver(ApplicationMaster.scala:426)
at org.apache.spark.deploy.yarn.ApplicationMaster.runExecutorLauncher(ApplicationMaster.scala:292)
...
2016-05-12 10:21:24,925 | INFO  | [Thread-1] | Unregistering ApplicationMaster with FAILED (diag
message: Uncaught exception: org.apache.spark.SparkException: Failed to connect to driver!) |
org.apache.spark.Logging$class.logInfo(Logging.scala:59)
```

   In Spark-client mode, the task Driver runs on a client node (usually a node outside the cluster). During the startup, the ApplicationMaster process is started in the cluster. After the process is started, information needs to be registered with the Driver process. The task can be continued only after the registration is successful. According to the ApplicationMaster log, the connection to the Driver fails, which causes the task failure.

Solution

Step 1 Check whether the IP address of the Driver process can be pinged.

Step 2 Start a SparkPI task. Information similar to the following is displayed on the console:

```
16/05/11 18:07:20 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://
sparkDriver@192.168.1.100:23662]
16/05/11 18:07:20 INFO Utils: Successfully started service 'sparkDriver' on port 23662.
```

Step 3 Run the **netstat - anp | grep 23662** command on the node (192.168.1.100 in **Step 2**) to check whether the port is enabled. The following information indicates that the port is enabled.

```
tcp        0      0  ip:port    :::*                    LISTEN       107274/java
tcp        0      0  ip:port   ip:port                    ESTABLISHED 107274/java
```

Step 4 Run the **telnet 192.168.1.100 23662** command on the node where ApplicationMaster is started to check whether the port can be connected. Perform this operation as both the **root** and **omm** users. If information similar to **Escape character is '^]'** is displayed, the connection is normal. If **connection refused** is displayed, the connection fails and the related port cannot be connected.

If the port is enabled but cannot be connected from other nodes, check the network configuration.

NOTE

The port (port 23662 in this example) is randomly selected each time. Therefore, you need to test the port enabled by the task.

----End

15.16.10 Failed to Connect to ResourceManager When a Spark Task Is Submitted

Symptom

The connection to ResourceManager is abnormal. As a result, Spark tasks fail to be submitted.

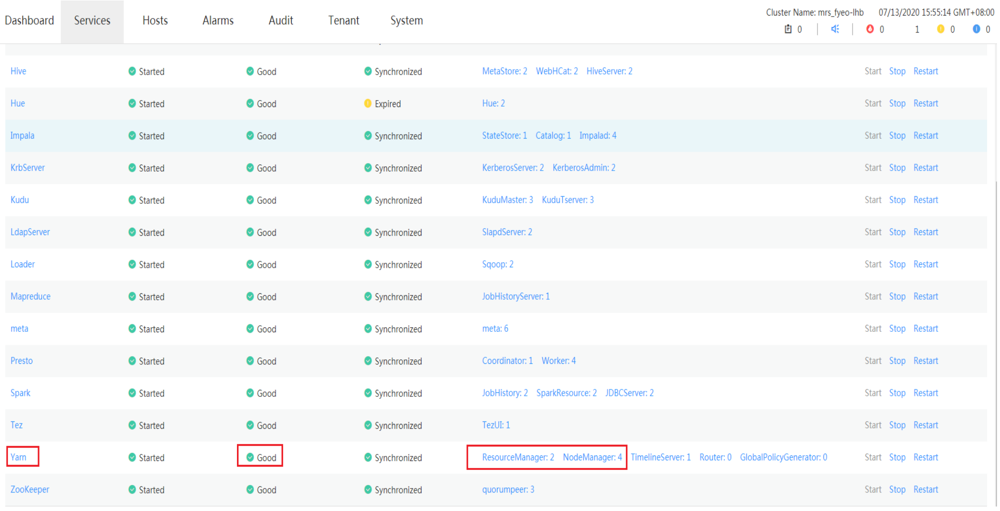
Cause Analysis

1. The following error information is displayed on the Driver, indicating that port 26004 connecting to the active and standby ResourceManager nodes is rejected:

```
15/08/19 18:36:16 INFO RetryInvocationHandler: Exception while invoking getClusterMetrics of class
ApplicationClientProtocolPBClientImpl over 33 after 1 fail over attempts. Trying to fail over after
sleeping for 17448ms.
 java.net.ConnectException: Call From ip0 to ip1:26004 failed on connection exception:
java.net.ConnectException: Connection refused.
INFO RetryInvocationHandler: Exception while invoking getClusterMetrics of class
ApplicationClientProtocolPBClientImpl over 32 after 2 fail over attempts. Trying to fail over after
sleeping for 16233ms.
 java.net.ConnectException: Call From ip0 to ip2:26004 failed on connection exception:
java.net.ConnectException: Connection refused;
```

2. On MRS Manager, check whether ResourceManager is running properly, as shown in **Figure 15-58**. If Yarn is faulty or an unknown exception occurs on a Yarn service instance, ResourceManager of the cluster may be abnormal.

Figure 15-58 Service status



Cluster Name: mrs_fyeo-lhb 07/13/2020 15:55:14 GMT+08:00

Hive	Started	Good	Synchronized	MetaStore: 2 WebHCat: 2 HiveServer: 2	Start Stop Restart
Hue	Started	Good	Expired	Hue: 2	Start Stop Restart
Impala	Started	Good	Synchronized	StateStore: 1 Catalog: 1 Impalad: 4	Start Stop Restart
KrbServer	Started	Good	Synchronized	KerberosServer: 2 KerberosAdmin: 2	Start Stop Restart
Kudu	Started	Good	Synchronized	KuduMaster: 3 KuduTserver: 3	Start Stop Restart
LdapServer	Started	Good	Synchronized	SlapdServer: 2	Start Stop Restart
Loader	Started	Good	Synchronized	Sqoop: 2	Start Stop Restart
Mapreduce	Started	Good	Synchronized	JobHistoryServer: 1	Start Stop Restart
meta	Started	Good	Synchronized	meta: 6	Start Stop Restart
Presto	Started	Good	Synchronized	Coordinator: 1 Worker: 4	Start Stop Restart
Spark	Started	Good	Synchronized	JobHistory: 2 SparkResource: 2 JDBCServer: 2	Start Stop Restart
Tez	Started	Good	Synchronized	TezUI: 1	Start Stop Restart
Yarn	Started	Good	Synchronized	ResourceManager: 2 NodeManager: 4 TimelineServer: 1 Router: 0 GlobalPolicyGenerator: 0	Start Stop Restart
ZooKeeper	Started	Good	Synchronized	quorumpeer: 3	Start Stop Restart

3. Check whether the client is the latest one in the cluster.

   Check whether the ResourceManager instance has been migrated in the cluster. (Uninstall a ResourceManager instance and add it back to other nodes.)

4. On MRS Manager, click **Audit** to view audit logs and check whether related operations are recorded.

   Run the **ping** command to check whether the IP address can be pinged.

Solution

- If ResourceManager is abnormal, see the Yarn-related sections to rectify the fault.
- If the client is not the latest, download the client again.
- If the IP address cannot be pinged, contact network management personnel to check the network.

15.16.11 DataArts Studio Failed to Schedule Spark Jobs

Issue

DataArts Studio fails to schedule jobs, and a message is displayed indicating that data in the **/thriftserver/active_thriftserver** directory cannot be read.

Symptom

DataArts Studio fails to schedule jobs, and the following error is reported indicating that data in the **/thriftserver/active_thriftserver** directory cannot be read:

Can not get JDBC Connection, due to KeeperErrorCode = NoNode for /thriftserver/active_thriftserver

Cause Analysis

When DataArts Studio submits a Spark job, Spark JDBC is invoked. Spark starts a ThriftServer process for the client to provide JDBC connections. During the startup, JDBCServer creates the **active_thriftserver** subdirectory in the **/thriftserver** directory of ZooKeeper, and registers related connection information. If the connection information cannot be read, the JDBC connection is abnormal.

Procedure

Check whether the ZooKeeper directory contains the target directory and registration information.

Step 1 Log in to any master node as user **root** and initialize environment variables.

source /opt/client/bigdata_env

Step 2 Run the **zkCli.sh -server 'ZookeeperIp:2181'** command to log in to ZooKeeper.

Step 3 Run the **ls /thriftserver** command to check whether the **active_thriftserver** directory exists.

- If the **active_thriftserver** directory exists, run the **get /thriftserver/active_thriftserver** command to check whether it contains the registered configuration information.
  - If yes, contact technical support.
  - If no, go to **Step 4**.
- If the **active_thriftserver** directory does not exist, go to **Step 4**.

Step 4 Log in to Manager and check whether the active/standby status of the Spark JDBCServer instance is unknown.

- If yes, go to **Step 5**.
- If no, contact O&M personnel.

Step 5 Restart the two JDBCServer instances. Check whether the status of the active and standby instances is normal and whether the target directory and data exist in ZooKeeper. If yes, the job is restored. If the instance status is not restored, contact technical support.

----End

15.16.12 Submission Status of the Spark Job API Is Error

Issue

After a Spark job is submitted using an API, the job status is displayed as **error**.

Issue Type

Job management

Symptom

After the log level in **/opt/client/Spark/spark/conf/log4j.properties** is changed and a job is submitted using API V1.1, the job status is displayed as error.

Cause Analysis

The executor monitors the job log output and determines the job execution result. After the execution result is changed to **error**, the output result cannot be detected. Therefore, the executor determines that the job status is abnormal after the job expires.

Procedure

Change the log level in the **/opt/client/Spark/spark/conf/log4j.properties** file to **info**.

Summary and Suggestions

You are advised to use the V2 API to submit jobs.

15.16.13 Alarm 43006 Is Repeatedly Generated in the Cluster

Issue

The alarm "ALM-43006 Heap Memory Usage of the JobHistory Process Exceeds the Threshold" is repeatedly generated in the cluster, and the setting according to the alarm reference is invalid.

Symptom

Alarm **ALM-43006 Heap Memory Usage of the JobHistory Process Exceeds the Threshold** is generated in the cluster. The same alarm is generated again a period of time after handling measures are taken.

Cause Analysis

The JobHistory memory leakage may occur. You need to install the corresponding patch to rectify the fault.

Procedure

- Increase the heap memory of the JobHistory process.
- If the heap memory has been increased, restart the JobHistory instance.

15.16.14 Failed to Create or Delete a Table in Spark Beeline

Issue

When the customer frequently creates or deletes a large number of users in Spark Beeline, some users occasionally fail to create or delete tables.

Symptom

The procedure for creating a table is as follows:

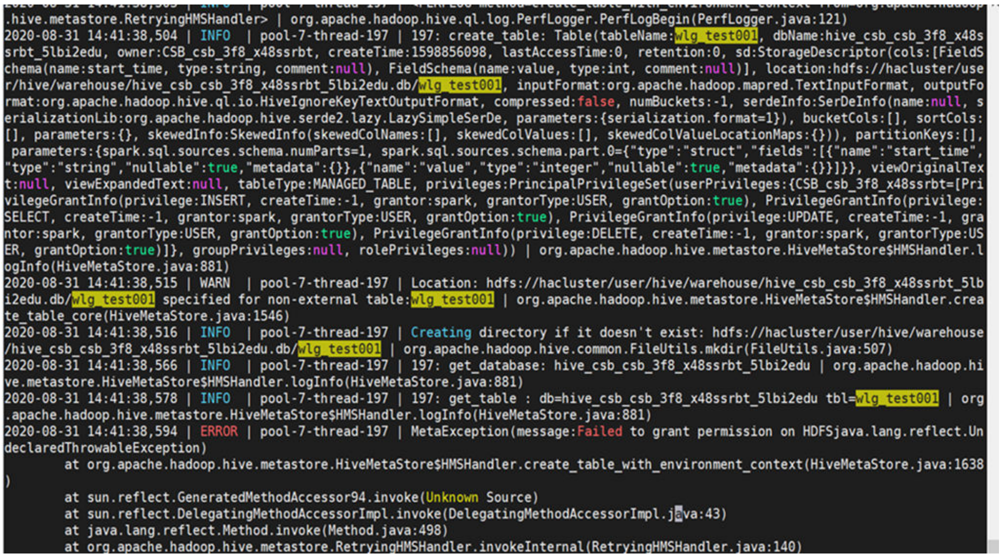
```
CREATE TABLE wlg_test001 (start_time STRING,value INT);
```

The following error message is displayed:

```
Error: org.apache.spark.sql.AnalysisException:
org.apache.hadoop.hive.ql.metadata.HiveException: MetaException(message:Failed to grant permission on
HDFSjava.lang.reflect.UndeclaredThrowableException); (state=,code=0)
```

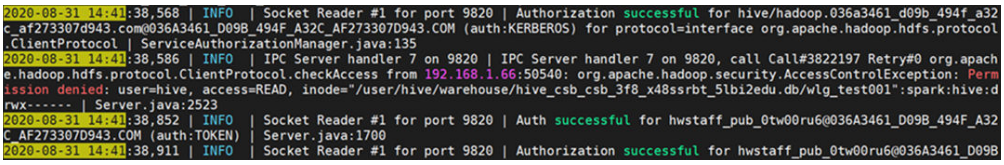
Cause Analysis

1. View metastore logs.



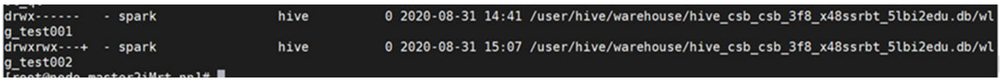
```
.hive.metastore.RetryingHMSHandler> | org.apache.hadoop.hive.ql.log.PerfLogger.PerfLogBegin(PerfLogger.java:121)
2020-08-31 14:41:38,504 | INFO  | pool-7-thread-197 | 197: create_table: Table(tableName:wlg_test001, dbName:hive_csb_csb_3f8_x48s
srbt_5lbi2edu, owner:CSB_csb_3f8_x48ssrbt, createTime:1598856098, lastAccessTime:0, retention:0, sd:StorageDescriptor(cols:[FieldS
chema(name:start_time, type:string, comment:null), FieldSchema(name:value, type:int, comment:null)], location:hdfs://hacluster/use
r/hive/warehouse/hive_csb_csb_3f8_x48ssrbt_5lbi2edu.db/wlg_test001, inputFormat:org.apache.hadoop.mapred.TextInputFormat, outputFo
rmat:org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat, compressed:false, numBuckets:-1, serdeInfo:SerDeInfo(name:null, s
erializationLib:org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe, parameters:{serialization.format=1}), bucketCols:[], sortCols:
[], parameters:{}, skewedInfo:SkewedInfo(skewedColNames:[], skewedColValues:[], skewedColValueLocationMaps:{})), partitionKeys:[],
 parameters:{spark.sql.sources.schema.numParts=1, spark.sql.sources.schema.part.0={"type":"struct","fields":[{"name":"start_time",
"type":"string","nullable":true,"metadata":{}},{"name":"value","type":"integer","nullable":true,"metadata":{}}]}}, viewOriginalTex
t:null, viewExpandedText:null, tableType:MANAGED_TABLE, privileges:PrincipalPrivilegeSet(userPrivileges:{CSB_csb_3f8_x48ssrbt=[Pri
vilegeGrantInfo(privilege:INSERT, createTime:-1, grantor:spark, grantorType:USER, grantOption:true), PrivilegeGrantInfo(privilege:
SELECT, createTime:-1, grantor:spark, grantorType:USER, grantOption:true), PrivilegeGrantInfo(privilege:UPDATE, createTime:-1, gra
ntor:spark, grantorType:USER, grantOption:true), PrivilegeGrantInfo(privilege:DELETE, createTime:-1, grantor:spark, grantorType:US
ER, grantOption:true)]}, groupPrivileges:null, rolePrivileges:null)) | org.apache.hadoop.hive.metastore.HiveMetaStore$HMSHandler.l
ogInfo(HiveMetaStore.java:881)
2020-08-31 14:41:38,515 | WARN  | pool-7-thread-197 | Location: hdfs://hacluster/user/hive/warehouse/hive_csb_csb_3f8_x48ssrbt_5lb
i2edu.db/wlg_test001 specified for non-external table:wlg_test001 | org.apache.hadoop.hive.metastore.HiveMetaStore$HMSHandler.crea
te_table_core(HiveMetaStore.java:1546)
2020-08-31 14:41:38,516 | INFO  | pool-7-thread-197 | Creating directory if it doesn't exist: hdfs://hacluster/user/hive/warehouse
/hive_csb_csb_3f8_x48ssrbt_5lbi2edu.db/wlg_test001 | org.apache.hadoop.hive.common.FileUtils.mkdir(FileUtils.java:507)
2020-08-31 14:41:38,566 | INFO  | pool-7-thread-197 | 197: get_database: hive_csb_csb_3f8_x48ssrbt_5lbi2edu | org.apache.hadoop.hi
ve.metastore.HiveMetaStore$HMSHandler.logInfo(HiveMetaStore.java:881)
2020-08-31 14:41:38,578 | INFO  | pool-7-thread-197 | 197: get_table : db=hive_csb_csb_3f8_x48ssrbt_5lbi2edu tbl=wlg_test001 | org
.apache.hadoop.hive.metastore.HiveMetaStore$HMSHandler.logInfo(HiveMetaStore.java:881)
2020-08-31 14:41:38,594 | ERROR | pool-7-thread-197 | MetaException(message:Failed to grant permission on HDFSjava.lang.reflect.Un
declaredThrowableException)
        at org.apache.hadoop.hive.metastore.HiveMetaStore$HMSHandler.create_table_with_environment_context(HiveMetaStore.java:1638
)
        at sun.reflect.GeneratedMethodAccessor94.invoke(Unknown Source)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.hadoop.hive.metastore.RetryingHMSHandler.invokeInternal(RetryingHMSHandler.java:140)
```

2. View HDFS logs.



```
2020-08-31 14:41:38,568 | INFO  | Socket Reader #1 for port 9820 | Authorization successful for hive/hadoop.036a3461_d09b_494f_a32
c_af273307d943.com@036A3461_D09B_494F_A32C_AF273307D943.COM (auth:KERBEROS) for protocol=interface org.apache.hadoop.hdfs.protocol
.ClientProtocol | ServiceAuthorizationManager.java:135
2020-08-31 14:41:38,586 | INFO  | IPC Server handler 7 on 9820 | IPC Server handler 7 on 9820, call Call#3822197 Retry#0 org.apach
e.hadoop.hdfs.protocol.ClientProtocol.checkAccess from 192.168.1.66:50540: org.apache.hadoop.security.AccessControlException: Perm
ission denied: user=hive, access=READ, inode="/user/hive/warehouse/hive_csb_csb_3f8_x48ssrbt_5lbi2edu.db/wlg_test001":spark:hive:d
rwx------ | Server.java:2523
2020-08-31 14:41:38,852 | INFO  | Socket Reader #1 for port 9820 | Auth successful for hwstaff_pub_0tw00ru6@036A3461_D09B_494F_A32
C_AF273307D943.COM (auth:TOKEN) | Server.java:1700
2020-08-31 14:41:38,911 | INFO  | Socket Reader #1 for port 9820 | Authorization successful for hwstaff_pub_0tw00ru6@036A3461_D09B
```

3. Compare permission (**test001** is a table created by a user in abnormal state, and **test002** is a table created by a user in normal state).



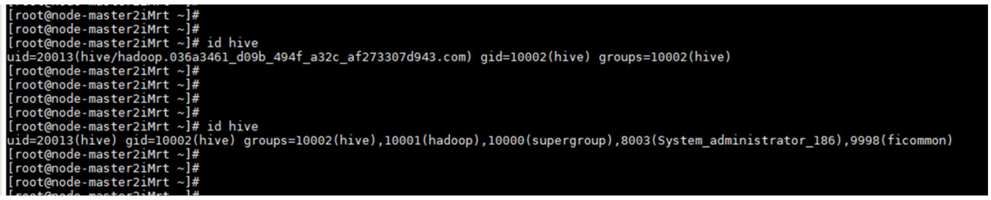
```
drwx------   - spark            hive         0 2020-08-31 14:41 /user/hive/warehouse/hive_csb_csb_3f8_x48ssrbt_5lbi2edu.db/wl
g_test001
drwxrwx---+  - spark            hive         0 2020-08-31 15:07 /user/hive/warehouse/hive_csb_csb_3f8_x48ssrbt_5lbi2edu.db/wl
g_test002
```

4. An error similar to the following is reported when a table is dropped:

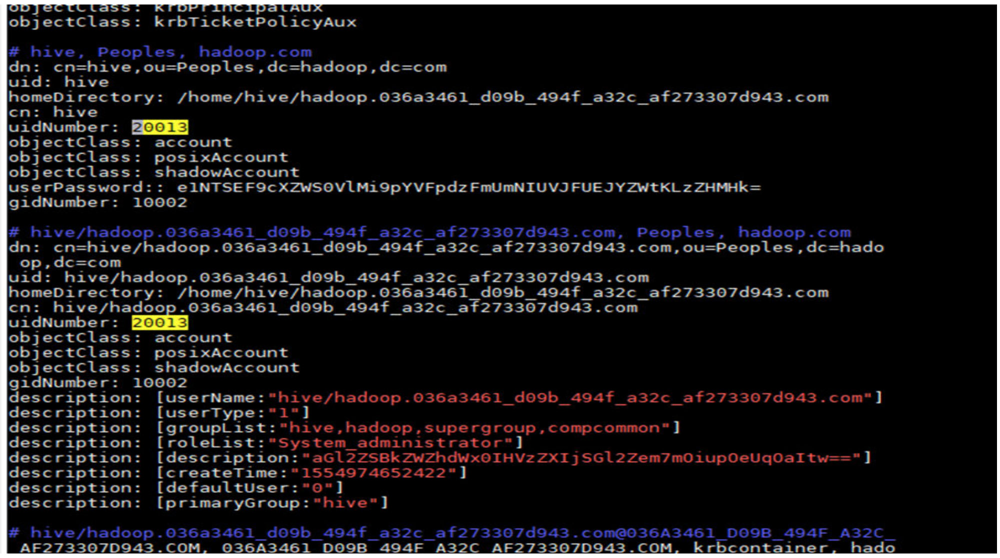
```
0: jdbc:hive2://192.168.1.42:10000/> drop table
dataplan_modela_csbch2;
Error: Error while compiling statement: FAILED:
SemanticException Unable to fetch table dataplan_modela_csbch2.
java.security.AccessControlException: Permission denied: user=CSB_csb_3f8_x48ssrbt,
access=READ,
inode="/user/hive/warehouse/hive_csb_csb_3f8_x48ssrbt_5lbi2edu.db/
dataplan_modela_csbch2":spark:hive:drwx------
```

5. Analyze the cause.

   The default user created during cluster creation uses the same UID, causing user disorder. This problem is triggered when a large number of users are created. As a result, the Hive user does not have the permission to create tables occasionally.



```
[root@node-master2iMrt ~]#
[root@node-master2iMrt ~]#
[root@node-master2iMrt ~]# id hive
uid=20013(hive/hadoop.036a3461_d09b_494f_a32c_af273307d943.com) gid=10002(hive) groups=10002(hive)
[root@node-master2iMrt ~]#
[root@node-master2iMrt ~]#
[root@node-master2iMrt ~]#
[root@node-master2iMrt ~]#
[root@node-master2iMrt ~]# id hive
uid=20013(hive) gid=10002(hive) groups=10002(hive),10001(hadoop),10000(supergroup),8003(System_administrator_186),9998(ficommon)
[root@node-master2iMrt ~]#
[root@node-master2iMrt ~]#
[root@node-master2iMrt ~]#
```



```
objectClass: krbPrincipalAux
objectClass: krbTicketPolicyAux

# hive, Peoples, hadoop.com
dn: cn=hive,ou=Peoples,dc=hadoop,dc=com
uid: hive
homeDirectory: /home/hive/hadoop.036a3461_d09b_494f_a32c_af273307d943.com
cn: hive
uidNumber: 20013
objectClass: account
objectClass: posixAccount
objectClass: shadowAccount
userPassword:: e1NTSEF9cXZWS0VlMi9pYVFpdzFmUmNIUVJFUEJYZWtKLzZHMHk=
gidNumber: 10002

# hive/hadoop.036a3461_d09b_494f_a32c_af273307d943.com, Peoples, hadoop.com
dn: cn=hive/hadoop.036a3461_d09b_494f_a32c_af273307d943.com,ou=Peoples,dc=hado
 op,dc=com
uid: hive/hadoop.036a3461_d09b_494f_a32c_af273307d943.com
homeDirectory: /home/hive/hadoop.036a3461_d09b_494f_a32c_af273307d943.com
cn: hive/hadoop.036a3461_d09b_494f_a32c_af273307d943.com
uidNumber: 20013
objectClass: account
objectClass: posixAccount
objectClass: shadowAccount
gidNumber: 10002
description: [userName:"hive/hadoop.036a3461_d09b_494f_a32c_af273307d943.com"]
description: [userType:"1"]
description: [groupList:"hive,hadoop,supergroup,compcommon"]
description: [roleList:"System_administrator"]
description: [description:"aGl2ZSBkZWZhdWx0IHVzZXIjSGl2Zem7mOiupOeUqOaItw=="]
description: [createTime:"1554974652422"]
description: [defaultUser:"0"]
description: [primaryGroup:"hive"]

# hive/hadoop.036a3461_d09b_494f_a32c_af273307d943.com@036A3461_D09B_494F_A32C_
 AF273307D943.COM, 036A3461_D09B_494F_A32C_AF273307D943.COM, krbcontainer, hado
```

Procedure

Restart the **sssd** process of the cluster.

Run the **service sssd restart** command as the **root** user to restart the **sssd** process and run the **ps -ef | grep sssd** command to check whether the **sssd** process is running properly.

In normal cases, the **/usr/sbin/sssd** process and three sub-processes **/usr/libexec/sssd/sssd_be**, **/usr/libexec/sssd/sssd_nss** and **/usr/libexec/sssd/sssd_pam** exist.

15.16.15 Failed to Connect to the Driver When a Node Outside the Cluster Submits a Spark Job to Yarn

Issue

When a node outside the cluster uses the client mode to submit a Spark task to Yarn, the task fails and an error message is displayed, indicating that the driver cannot be connected.

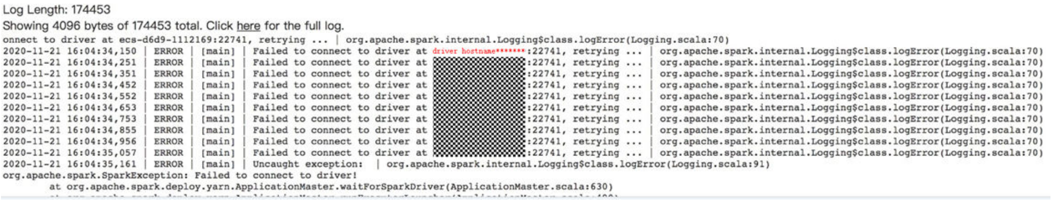
Symptom

Nodes outside the cluster can communicate with each node in the cluster. When a node outside the cluster submits a Spark task to Yarn in client mode, the task fails and an error message is displayed, indicating that the driver cannot be connected.

Cause Analysis

When a Spark task is submitted in the client mode, the driver process of Spark is on the client side, and the executor needs to interact with the driver to run the job.

If the NodeManager fails to connect to the node where the client is located, the following error is reported:



```
Log Length: 174453
Showing 4096 bytes of 174453 total. Click here for the full log.
onnect to driver at ecs-d6d9-1112169:22741, retrying ... | org.apache.spark.internal.Logging$class.logError(Logging.scala:70)
2020-11-21 16:04:34,150 | ERROR | [main] | Failed to connect to driver at driver hostname*******:22741, retrying ... | org.apache.spark.internal.Logging$class.logError(Logging.scala:70)
2020-11-21 16:04:34,251 | ERROR | [main] | Failed to connect to driver at [redacted]:22741, retrying ... | org.apache.spark.internal.Logging$class.logError(Logging.scala:70)
2020-11-21 16:04:34,351 | ERROR | [main] | Failed to connect to driver at :22741, retrying ... | org.apache.spark.internal.Logging$class.logError(Logging.scala:70)
2020-11-21 16:04:34,452 | ERROR | [main] | Failed to connect to driver at :22741, retrying ... | org.apache.spark.internal.Logging$class.logError(Logging.scala:70)
2020-11-21 16:04:34,552 | ERROR | [main] | Failed to connect to driver at :22741, retrying ... | org.apache.spark.internal.Logging$class.logError(Logging.scala:70)
2020-11-21 16:04:34,653 | ERROR | [main] | Failed to connect to driver at :22741, retrying ... | org.apache.spark.internal.Logging$class.logError(Logging.scala:70)
2020-11-21 16:04:34,753 | ERROR | [main] | Failed to connect to driver at :22741, retrying ... | org.apache.spark.internal.Logging$class.logError(Logging.scala:70)
2020-11-21 16:04:34,855 | ERROR | [main] | Failed to connect to driver at :22741, retrying ... | org.apache.spark.internal.Logging$class.logError(Logging.scala:70)
2020-11-21 16:04:34,956 | ERROR | [main] | Failed to connect to driver at :22741, retrying ... | org.apache.spark.internal.Logging$class.logError(Logging.scala:70)
2020-11-21 16:04:35,057 | ERROR | [main] | Failed to connect to driver at :22741, retrying ... | org.apache.spark.internal.Logging$class.logError(Logging.scala:70)
2020-11-21 16:04:35,161 | ERROR | [main] | Uncaught exception:  | org.apache.spark.internal.Logging$class.logError(Logging.scala:91)
org.apache.spark.SparkException: Failed to connect to driver!
        at org.apache.spark.deploy.yarn.ApplicationMaster.waitForSparkDriver(ApplicationMaster.scala:630)
```

Procedure

Specify the IP address of the driver in the Spark configuration of the client.

Add **spark.driver.host=driverIP** to *<Client installation path>***/Spark/spark/conf/spark-defaults.conf** and run the Spark task again.

Summary and Suggestions

You are advised to submit jobs in cluster mode.

15.16.16 Large Number of Shuffle Results Are Lost During Spark Task Execution

Issue

Spark tasks fail to be executed. The task log shows that a large number of **shuffle** files are lost.

Symptom

Spark tasks fail to be executed. The task log shows that a large number of **shuffle** files are lost.

Cause Analysis

When Spark is running, the **shuffle** file generated temporarily is stored in the temporary directory of the executor for later use.

When an executor exits abnormally, NodeManager deletes the temporary directory of the container where the executor is located. When other executors apply for the shuffle result of the executor, a message is displayed indicating that the file cannot be found.

Therefore, you need to check whether the executor exits abnormally. You can check whether there are executors in the **dead** state on the executors tab page on the Spark task page and view the executor logs of each **dead** state, determine the cause of abnormal exit. Some executors may exit because the **shuffle** file cannot be found. You need to find the earliest executor that exits abnormally.

Common abnormal exit causes:

- OOM occurs on the executor.
- Multiple tasks fail when the executor is running.
- The node where the executor is located is cleared.

Procedure

Adjust or modify the task parameters or code based on the actual cause of the abnormal exit of the executor, and run the Spark task again.

15.16.17 Disk Space Is Insufficient Due to Long-Term Running of JDBCServer

Issue

When the JDBCServer service connected to Spark submits a spark-sql task to the Yarn cluster, the data disk of the Core node is fully occupied after the task runs for a period of time.

Symptom

When the JDBCServer service of a customer connected to Spark submits a spark-sql task to the Yarn cluster, the data disk of the Core node is fully occupied after the task runs for a period of time.

After checking the disk usage in the background, it is found that there are too many APP temporary files (files generated by shuffle) of the JDBCServer service, and the files are not cleared, occupying a large amount of memory.

Cause Analysis

After checking the directories that contain a large number of files on the Core node, it is found that most of the directories are similar to **blockmgr-033707b6-fbbb-45b4-8e3a-128c9bcfa4bf**, which stores temporary shuffle files generated during computing.

The dynamic resource allocation function of Spark is enabled on JDBCServer, and shuffle is hosted by NodeManager. NodeManager only manages these files based on the running period of the application, and does not check whether the container where a single executor is located exists. Therefore, the temporary files are deleted only when the app is stopped. When a task runs for a long time, a large number of temporary files occupy a large amount of disk space.

Procedure

Start a scheduled task to delete shuffle files that have been stored for a specified period of time. For example, delete shuffle files that have been stored for more than 6 hours each hour.

Step 1 Create the **clean_appcache.sh** script. If there are multiple data disks, change the value of **data1** in **BASE_LOC** based on the actual situation.

- Security cluster

```
#!/bin/bash
BASE_LOC=/srv/BigData/hadoop/data1/nm/localdir/usercache/spark/appcache/application_*/
blockmgr*
find $BASE_LOC/ -mmin +360 -exec rmdir {} \;
find $BASE_LOC/ -mmin +360 -exec rm {} \;
```

- Common cluster

```
#!/bin/bash
BASE_LOC=/srv/BigData/hadoop/data1/nm/localdir/usercache/omm/appcache/application_*/
```

```
blockmgr*
find $BASE_LOC/ -mmin +360 -exec rmdir {} \;
find $BASE_LOC/ -mmin +360 -exec rm {} \;
```

Step 2 Run the following commands to change the permission to the script:

chmod 755 clean_appcache.sh

Step 3 Add a scheduled task to start the clearance script. Change the script path to the actual path.

Run the **crontab -l** command to view the scheduled task.

Run the **crontab -e** command to edit the scheduled task.

```
0 * * * * sh /root/clean_appcache.sh > /dev/null 2>&1
```

----End

15.16.18 Failed to Load Data to a Hive Table Across File Systems by Running SQL Statements Using Spark Shell

Issue

When the **spark-shell** command is used to execute SQL statements or the **spark-submit** command is used to submit Spark tasks, the **load** command of SQL statements exists, and the source data and target table are not stored in the same file system. An error is reported when the MapReduce task is started in the preceding two modes.

Cause Analysis

When the **load** command is used to import data to the Hive table across file systems (for example, the original data is stored in the HDFS but the Hive table data is stored in the OBS), and the file length is greater than the threshold (32 MB by default). In this case, the MapReduce job that uses DistCp is triggered to migrate data. The MapReduce task configuration is directly extracted from the Spark task configuration. However, the **net.topology.node.switch.mapping.impl** configuration item of the Spark task does not retain the default value of the Hadoop. Therefore, the JAR package of the Spark needs to be used. As a result, the MapReduce reports an error indicating that the class cannot be found.

Procedure

Solution 1:

If the file size is small, set the default file size to a value greater than the maximum file size. For example, if the maximum file size is 95 MB, run the following command:

```
hive.exec.copyfile.maxsize=104857600
```

Solution 2:

If the file size is large, use DistCp to improve the data migration efficiency. Add the following parameters when starting the Spark task:

```
--conf spark.hadoop.net.topology.node.switch.mapping.impl=org.apache.hadoop.net.ScriptBasedMapping
```

15.16.19 Spark Task Submission Failure

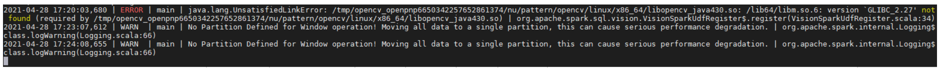
Symptom

- A Spark task fails to be submitted.
- Spark displays a message indicating that the Yarn JAR package cannot be obtained.
- A file is submitted for multiple times.

Cause Analysis

- Symptom 1:

  The most common cause for task submission failure is authentication failure.

```
2021-04-28 17:20:03,680 | ERROR | main | java.lang.UnsatisfiedLinkError: /tmp/opencv_openpnp6650342257652861374/nu/pattern/opencv/linux/x86_64/libopencv_java430.so: /lib64/libm.so.6: version `GLIBC_2.27' not
 found (required by /tmp/opencv_openpnp6650342257652861374/nu/pattern/opencv/linux/x86_64/libopencv_java430.so) | org.apache.spark.sql.vision.VisionSparkUdfRegister$.register(VisionSparkUdfRegister.scala:34)
2021-04-28 17:23:07,612 | WARN  | main | No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation. | org.apache.spark.internal.Logging$
class.logWarning(Logging.scala:66)
2021-04-28 17:24:08,655 | WARN  | main | No Partition Defined for Window operation! Moving all data to a single partition, this can cause serious performance degradation. | org.apache.spark.internal.Logging$
class.logWarning(Logging.scala:66)
```

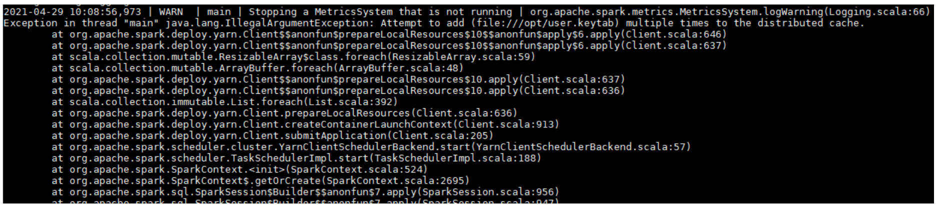
  The parameter settings may be incorrect.

- Symptom 2:

  By default, the cluster adds the Hadoop JAR package of the analysis node to the classpath of the task. If the system displays a message indicating that Yarn packages cannot be found, the Hadoop configuration is not set.

- Symptom 3:

  The common scenario is as follows: The **--files** option is used to upload the **user.keytab** file, and then the **--keytab** option is used to specify the same file. As a result, the same file is uploaded for multiple times.

```
2021-04-29 10:08:56,973 | WARN  | main | Stopping a MetricsSystem that is not running | org.apache.spark.metrics.MetricsSystem.logWarning(Logging.scala:66)
Exception in thread "main" java.lang.IllegalArgumentException: Attempt to add (file:///opt/user.keytab) multiple times to the distributed cache.
        at org.apache.spark.deploy.yarn.Client$$anonfun$prepareLocalResources$10$$anonfun$apply$6.apply(Client.scala:646)
        at org.apache.spark.deploy.yarn.Client$$anonfun$prepareLocalResources$10$$anonfun$apply$6.apply(Client.scala:637)
        at scala.collection.mutable.ResizableArray$class.foreach(ResizableArray.scala:59)
        at scala.collection.mutable.ArrayBuffer.foreach(ArrayBuffer.scala:48)
        at org.apache.spark.deploy.yarn.Client$$anonfun$prepareLocalResources$10.apply(Client.scala:637)
        at org.apache.spark.deploy.yarn.Client$$anonfun$prepareLocalResources$10.apply(Client.scala:636)
        at scala.collection.immutable.List.foreach(List.scala:392)
        at org.apache.spark.deploy.yarn.Client.prepareLocalResources(Client.scala:636)
        at org.apache.spark.deploy.yarn.Client.createContainerLaunchContext(Client.scala:913)
        at org.apache.spark.deploy.yarn.Client.submitApplication(Client.scala:205)
        at org.apache.spark.scheduler.cluster.YarnClientSchedulerBackend.start(YarnClientSchedulerBackend.scala:57)
        at org.apache.spark.scheduler.TaskSchedulerImpl.start(TaskSchedulerImpl.scala:188)
        at org.apache.spark.SparkContext.<init>(SparkContext.scala:524)
        at org.apache.spark.SparkContext$.getOrCreate(SparkContext.scala:2695)
        at org.apache.spark.sql.SparkSession$Builder$$anonfun$7.apply(SparkSession.scala:956)
        at org.apache.spark.sql.SparkSession$Builder$$anonfun$7.apply(SparkSession.scala:947)
```

Procedure

- Symptom 1:

  Run **kinit** *[user]* again and modify the corresponding configuration items.

- Symptom 2:

  Check that the Hadoop configuration items are correct and the **core-site.xml**, **hdfs-site.xml**, **yarn-site.xml**, and **mapred-site.xml** configuration files in the **conf** directory of Spark are correct.

- Symptom 3:

  Copy a new **user.keytab** file, for example:

  cp user.keytab user2.keytab

  spark-submit --master yarn --files user.keytab --keytab user2.keytab ......

15.16.20 Spark Task Execution Failure

Symptom

- An executor out of memory (OOM) error occurs.
- The information about the failed task shows that the failure cause is "lost task xxx."

Cause Analysis

- Symptom 1: The data volume is too large or too many tasks are running on the same executor at the same time.
- Symptom 2: Some tasks fail to be executed. When the error is reported, determine the node where the lost task is running. Generally, the error is caused by the abnormal exit of the lost task.

Procedure

- Symptom 1:
  - If the data volume is too large, adjust the memory size of the executor and use **--executor-memory** to specify the memory size.
  - If too many tasks are running at the same time, check the number of vcores specified by **--executor-cores**.
- Symptom 2: Locate the cause in the corresponding task log. If an OOM error occurs, see the solutions to symptom 1.

15.16.21 JDBCServer Connection Failure

Symptom

- The ha-cluster cannot be identified (unknowHost or port required).
- Failed to connect to JDBCServer.

Cause Analysis

- Symptom 1: The **spark-beeline** command is used to connect to JDBCServer. JDBCServer in versions earlier than MRS_3.0 adopts HA mode. Therefore, a specific URL and the JAR package provided by MRS Spark is required to connect to JDBCServer.
- Symptom 2: The JDBCServer service is not running properly or port listening is abnormal.

Procedure

- Symptom 1: Use a specific URL and the JAR package provided by MRS Spark to connect to JDBCServer.
- Symptom 2: Check that the JDBCServer service is running properly and port listening is normal, and try again.

15.16.22 Failed to View Spark Task Logs

Symptom

- A user fails to view logs when a task is running.
- A user fails to view logs when a task is complete.

Cause Analysis

- Symptom 1: The MapReduce component is abnormal.
- Symptom 2:
  - The JobHistory service of Spark is abnormal.
  - The log size is too large, and NodeManager times out during log aggregation.
  - The permission on the HDFS log storage directory (**/tmp/logs/Username/logs** by default) is abnormal.
  - Logs have been deleted. By default, Spark JobHistory stores event logs for seven days (specified by **spark.history.fs.cleaner.maxAge**). MapReduce stores task logs for 15 days (specified by **mapreduce.jobhistory.max-age-ms**).
  - If the task cannot be found on the Yarn page, it may have been cleared by Yarn. By default, Yarn stores 10,000 historical tasks (specified by **yarn.resourcemanager.max-completed-applications**).

Procedure

- Symptom 1: Check whether the MapReduce component is running properly. If it is abnormal, restart it. If the fault persists, check the JobhistoryServer log file in the background.
- Symptom 2: Perform the following checks in sequence:
  a. Check whether JobHistory of Spark is running properly.
  b. On the app details page of Yarn, check whether the log file is too large. If log aggregation fails, the value of **Log Aggregation Status** should be **Failed** or **Timeout**.
  c. Check whether the permission on the corresponding directory is normal.
  d. Check whether the corresponding **appid** file exists in the directory. In MRS 3.x or later, the event log files are stored in the **hdfs://hacluster/spark2xJobHistory2x** directory. In versions earlier than MRS 3.x, the event log files are stored in the **hdfs://hacluster/sparkJobHistory** directory. The task run logs are stored in the **hdfs://hacluster/tmp/logs/*Username*/logs** directory.
  e. Check whether **appid** or the current job ID exceeds the maximum value in the historical records.

15.16.23 Authentication Fails When Spark Connects to Other Services

Symptom

- When Spark connects to HBase, an authentication failure message is displayed or the HBase table cannot be connected.
- When Spark connects to HBase, a message is displayed indicating that the JAR package cannot be found.

Cause Analysis

- Symptom 1: HBase does not obtain the authentication information of the current task. As a result, the authentication fails when HBase is connected, and the corresponding data cannot be read
- Symptom 2: By default, Spark does not load the HBase JAR package. You need to use **--jars** to add the JAR package to the task.

Procedure

- Symptom 1: Enable the HBase authentication function by running the **spark.yarn.security.credentials.hbase.enabled=true** command. However, do not replace **hbase-site.xml** on the Spark client with **hbase-site.xml** on the HBase client because they are not completely consistent.
- Symptom 2: Use **--jars** to upload the HBase JAR package.

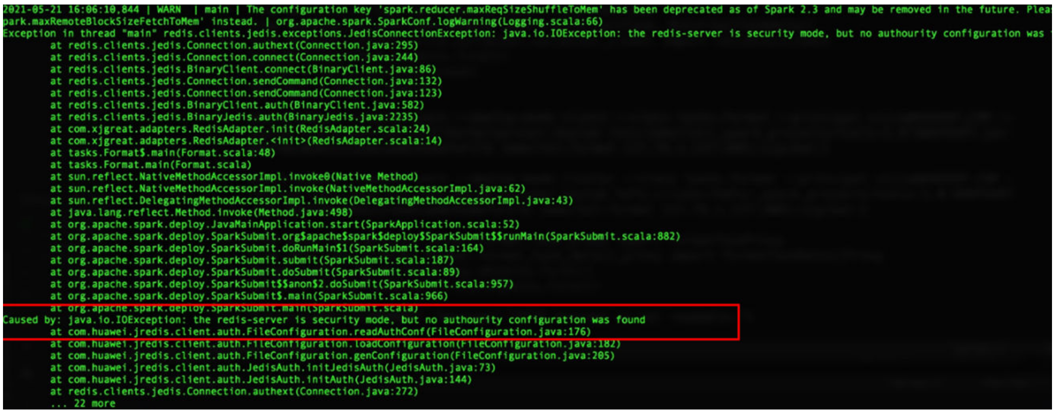
15.16.24 An Error Occurs When Spark Connects to Redis

Issue

An error occurs when the Spark component of the MRS 3.x security cluster is used to access Redis.

Symptom

When Spark of the MRS 3.0 security cluster is used to access Redis, the following error message is displayed.



```
2021-05-21 16:06:10,844 | WARN  | main | The configuration key 'spark.reducer.maxReqSizeShuffleToMem' has been deprecated as of Spark 2.3 and may be removed in the future. Plea
park.maxRemoteBlockSizeFetchToMem' instead. | org.apache.spark.SparkConf.logWarning(Logging.scala:66)
Exception in thread "main" redis.clients.jedis.exceptions.JedisConnectionException: java.io.IOException: the redis-server is security mode, but no authourity configuration was
        at redis.clients.jedis.Connection.authext(Connection.java:295)
        at redis.clients.jedis.Connection.connect(Connection.java:244)
        at redis.clients.jedis.BinaryClient.connect(BinaryClient.java:86)
        at redis.clients.jedis.Connection.sendCommand(Connection.java:132)
        at redis.clients.jedis.Connection.sendCommand(Connection.java:123)
        at redis.clients.jedis.BinaryClient.auth(BinaryClient.java:582)
        at redis.clients.jedis.BinaryJedis.auth(BinaryJedis.java:2235)
        at com.xjgreat.adapters.RedisAdapter.init(RedisAdapter.scala:24)
        at com.xjgreat.adapters.RedisAdapter.<init>(RedisAdapter.scala:14)
        at tasks.Format$.main(Format.scala:48)
        at tasks.Format.main(Format.scala)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.spark.deploy.JavaMainApplication.start(SparkApplication.scala:52)
        at org.apache.spark.deploy.SparkSubmit.org$apache$spark$deploy$SparkSubmit$$runMain(SparkSubmit.scala:882)
        at org.apache.spark.deploy.SparkSubmit.doRunMain$1(SparkSubmit.scala:164)
        at org.apache.spark.deploy.SparkSubmit.submit(SparkSubmit.scala:187)
        at org.apache.spark.deploy.SparkSubmit.doSubmit(SparkSubmit.scala:89)
        at org.apache.spark.deploy.SparkSubmit$$anon$2.doSubmit(SparkSubmit.scala:957)
        at org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:966)
        at org.apache.spark.deploy.SparkSubmit.main(SparkSubmit.scala)
Caused by: java.io.IOException: the redis-server is security mode, but no authourity configuration was found
        at com.huawei.jredis.client.auth.FileConfiguration.readAuthConf(FileConfiguration.java:176)
        at com.huawei.jredis.client.auth.FileConfiguration.loadConfiguration(FileConfiguration.java:182)
        at com.huawei.jredis.client.auth.FileConfiguration.genConfiguration(FileConfiguration.java:205)
        at com.huawei.jredis.client.auth.JedisAuth.initJedisAuth(JedisAuth.java:73)
        at com.huawei.jredis.client.auth.JedisAuth.initAuth(JedisAuth.java:144)
        at redis.clients.jedis.Connection.authext(Connection.java:272)
        ... 22 more
```

Cause Analysis

The **jars** directory of Spark contains a **jredisclient-xxx.jar** package provided by the MRS cluster. This package is loaded when a Spark task connects to Redis, thereby causing this error. You can manually remove this package to rectify the fault.

Procedure

Step 1 Delete JAR packages from the Spark client.

cd $SPARK_HOME/jars

mv jredisclient-*.jar /tmp

Step 2 Delete JAR packages from the Spark server.

Log in to the nodes (generally two) where SparkResource2x is located.

mkdir /tmp/SparkResource2x

cd /opt/Bigdata/FusionInsight_Current/1_*_SparkResource2x/install/spark/jars/

mv jredisclient-*.jar /tmp/SparkResource2x

Step 3 Delete the **jredisclient** file from the HDFS.

1. Check configuration item **spark.yarn.archive** in the **$SPARK_HOME/conf/spark-defaults.conf** file to obtain the address of the **spark-archive-2x.zip** package.

   cat $SPARK_HOME/conf/spark-defaults.conf | grep "spark.yarn.archive"

2. Download the **spark-archive-2x.zip** package. (This section uses MRS 3.0.5 as an example. Modify the command based on the actual cluster version.)

   cd /opt

   mkdir sparkTmp

   cd sparkTmp

   hdfs dfs -get hdfs://hacluster/user/spark2x/jars/8.0.2.1/spark-archive-2x.zip

3. Decompress **spark-archive-2x.zip** and remove the package file.

   unzip spark-archive-2x.zip

   rm -f spark-archive-2x.zip

4. Remove the **jredisclient** package.

   rm -f jredisclient-*.jar

5. Compress the **spark-archive-2x.zip** package again.

   zip spark-archive-2x.zip ./*

6. Back up the original package from the HDFS to **tmp** and upload the newly compressed package to the HDFS.

   hdfs dfs -mv hdfs://hacluster/user/spark2x/jars/8.0.2.1/spark-archive-2x.zip /tmp

   hdfs dfs -put spark-archive-2x.zip hdfs://hacluster/user/spark2x/jars/8.0.2.1/spark-archive-2x.zip

7. Restart the JDBCServer service to prevent JDBCServer exceptions. The **jredisclient** file has been deleted from the **spark-archive-2x.zip** package.
8. Delete temporary files.

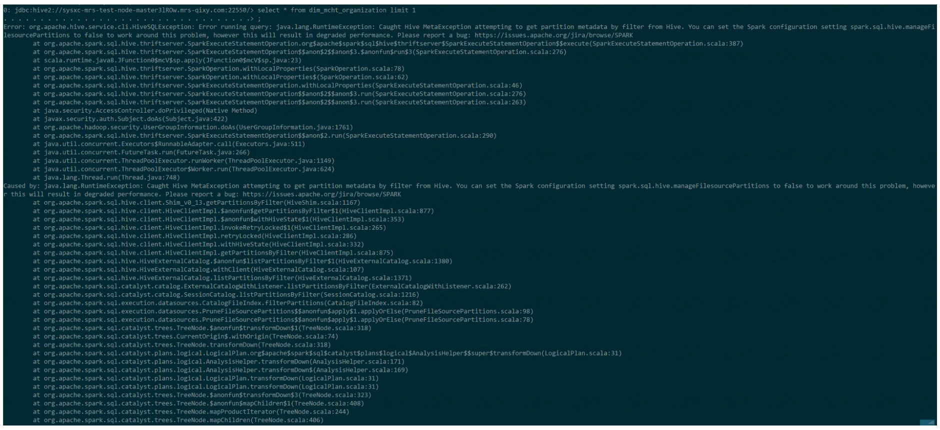
   rm -rf /opt/sparkTmp

----End

15.16.25 An Error Is Reported When spark-beeline Is Used to Query a Hive View

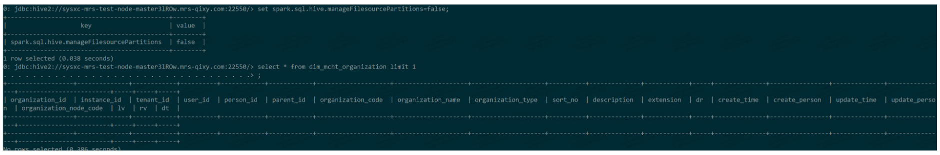
Issue

In MRS 3.1.2, an error is reported when spark-beeline is used to query a Hive view. The error information is as follows.



```
0: jdbc:hive2://sysxc-mrs-test-node-master31ROw.mrs-qixy.com:22550/> select * from dim_mcht_organization limit 1
. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .> ;
Error: org.apache.hive.service.cli.HiveSQLException: Error running query: java.lang.RuntimeException: Caught Hive MetaException attempting to get partition metadata by filter from Hive. You can set the Spark configuration setting spark.sql.hive.manageFi
lesourcePartitions to false to work around this problem, however this will result in degraded performance. Please report a bug: https://issues.apache.org/jira/browse/SPARK
        at org.apache.spark.sql.hive.thriftserver.SparkExecuteStatementOperation.org$apache$spark$sql$hive$thriftserver$SparkExecuteStatementOperation$$execute(SparkExecuteStatementOperation.scala:387)
        at org.apache.spark.sql.hive.thriftserver.SparkExecuteStatementOperation$$anon$2$$anon$3.$anonfun$run$3(SparkExecuteStatementOperation.scala:276)
        at scala.runtime.java8.JFunction0$mcV$sp.apply(JFunction0$mcV$sp.java:23)
        at org.apache.spark.sql.hive.thriftserver.SparkOperation.withLocalProperties(SparkOperation.scala:78)
        at org.apache.spark.sql.hive.thriftserver.SparkOperation.withLocalProperties$(SparkOperation.scala:62)
        at org.apache.spark.sql.hive.thriftserver.SparkExecuteStatementOperation.withLocalProperties(SparkExecuteStatementOperation.scala:46)
        at org.apache.spark.sql.hive.thriftserver.SparkExecuteStatementOperation$$anon$2$$anon$3.run(SparkExecuteStatementOperation.scala:276)
        at org.apache.spark.sql.hive.thriftserver.SparkExecuteStatementOperation$$anon$2$$anon$3.run(SparkExecuteStatementOperation.scala:263)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1761)
        at org.apache.spark.sql.hive.thriftserver.SparkExecuteStatementOperation$$anon$2.run(SparkExecuteStatementOperation.scala:290)
        at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:511)
        at java.util.concurrent.FutureTask.run(FutureTask.java:266)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
        at java.lang.Thread.run(Thread.java:748)
Caused by: java.lang.RuntimeException: Caught Hive MetaException attempting to get partition metadata by filter from Hive. You can set the Spark configuration setting spark.sql.hive.manageFilesourcePartitions to false to work around this problem, howeve
r this will result in degraded performance. Please report a bug: https://issues.apache.org/jira/browse/SPARK
        at org.apache.spark.sql.hive.client.Shim_v0_13.getPartitionsByFilter(HiveShim.scala:1167)
        at org.apache.spark.sql.hive.client.HiveClientImpl.$anonfun$getPartitionsByFilter$1(HiveClientImpl.scala:877)
        at org.apache.spark.sql.hive.client.HiveClientImpl.$anonfun$withHiveState$1(HiveClientImpl.scala:353)
        at org.apache.spark.sql.hive.client.HiveClientImpl.invokeRetryLocked$1(HiveClientImpl.scala:265)
        at org.apache.spark.sql.hive.client.HiveClientImpl.retryLocked(HiveClientImpl.scala:286)
        at org.apache.spark.sql.hive.client.HiveClientImpl.withHiveState(HiveClientImpl.scala:332)
        at org.apache.spark.sql.hive.client.HiveClientImpl.getPartitionsByFilter(HiveClientImpl.scala:875)
        at org.apache.spark.sql.hive.HiveExternalCatalog.$anonfun$listPartitionsByFilter$1(HiveExternalCatalog.scala:1380)
        at org.apache.spark.sql.hive.HiveExternalCatalog.withClient(HiveExternalCatalog.scala:107)
        at org.apache.spark.sql.hive.HiveExternalCatalog.listPartitionsByFilter(HiveExternalCatalog.scala:1371)
        at org.apache.spark.sql.catalyst.catalog.ExternalCatalogWithListener.listPartitionsByFilter(ExternalCatalogWithListener.scala:262)
        at org.apache.spark.sql.catalyst.catalog.SessionCatalog.listPartitionsByFilter(SessionCatalog.scala:1216)
        at org.apache.spark.sql.execution.datasources.CatalogFileIndex.filterPartitions(CatalogFileIndex.scala:82)
        at org.apache.spark.sql.execution.datasources.PruneFileSourcePartitions$$anonfun$apply$1.applyOrElse(PruneFileSourcePartitions.scala:98)
        at org.apache.spark.sql.execution.datasources.PruneFileSourcePartitions$$anonfun$apply$1.applyOrElse(PruneFileSourcePartitions.scala:78)
        at org.apache.spark.sql.catalyst.trees.TreeNode.$anonfun$transformDown$1(TreeNode.scala:318)
        at org.apache.spark.sql.catalyst.trees.CurrentOrigin$.withOrigin(TreeNode.scala:74)
        at org.apache.spark.sql.catalyst.trees.TreeNode.transformDown(TreeNode.scala:318)
        at org.apache.spark.sql.catalyst.plans.logical.LogicalPlan.org$apache$spark$sql$catalyst$plans$logical$AnalysisHelper$$super$transformDown(LogicalPlan.scala:31)
        at org.apache.spark.sql.catalyst.plans.logical.AnalysisHelper.transformDown(AnalysisHelper.scala:171)
        at org.apache.spark.sql.catalyst.plans.logical.AnalysisHelper.transformDown$(AnalysisHelper.scala:169)
        at org.apache.spark.sql.catalyst.plans.logical.LogicalPlan.transformDown(LogicalPlan.scala:31)
        at org.apache.spark.sql.catalyst.plans.logical.LogicalPlan.transformDown(LogicalPlan.scala:31)
        at org.apache.spark.sql.catalyst.trees.TreeNode.$anonfun$transformDown$3(TreeNode.scala:323)
        at org.apache.spark.sql.catalyst.trees.TreeNode.$anonfun$mapChildren$1(TreeNode.scala:408)
        at org.apache.spark.sql.catalyst.trees.TreeNode.mapProductIterator(TreeNode.scala:244)
        at org.apache.spark.sql.catalyst.trees.TreeNode.mapChildren(TreeNode.scala:406)
```

After **spark.sql.hive.manageFilesourcePartitions=false** is set as prompted, no data can be found. However, data can be queried in Hive.



```
0: jdbc:hive2://sysxc-mrs-test-node-master31ROw.mrs-qixy.com:22550/> set spark.sql.hive.manageFilesourcePartitions=false;
+---------------------------------------------+--------+
|                    key                      | value  |
+---------------------------------------------+--------+
| spark.sql.hive.manageFilesourcePartitions   | false  |
+---------------------------------------------+--------+
1 row selected (0.038 seconds)
0: jdbc:hive2://sysxc-mrs-test-node-master31ROw.mrs-qixy.com:22550/> select * from dim_mcht_organization limit 1
. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .> ;
+------------------+--------------+------------+----------+------------+------------+--------------------+--------------------+--------------------+----------+--------------+------------+-----+--------------+----------------+--------------+------------
---+-------------------------+-----+-----+-----+
| organization_id  | instance_id  | tenant_id  | user_id  | person_id  | parent_id  | organization_code  | organization_name  | organization_type  | sort_no  | description  | extension  | dr  | create_time  | create_person  | update_time  | update_perso
n  | organization_node_code  | lv  | rv  | dt  |
+------------------+--------------+------------+----------+------------+------------+--------------------+--------------------+--------------------+----------+--------------+------------+-----+--------------+----------------+--------------+------------
---+-------------------------+-----+-----+-----+
+------------------+--------------+------------+----------+------------+------------+--------------------+--------------------+--------------------+----------+--------------+------------+-----+--------------+----------------+--------------+------------
---+-------------------------+-----+-----+-----+
No rows selected (0.386 seconds)
```

Cause Analysis

Failed to convert the format. Spark SQL uses its built-in Metastore instead of Hive Metastore. As a result, metadata fails to be read. Spark SQL reads data in Parquet format, but Hive reads data in ORC format by default.

Procedure

Step 1 Log in to the Spark client node and run the following commands to access Spark SQL:

cd *Client installation directory*

source bigdata_env

source Spark2x/component_env

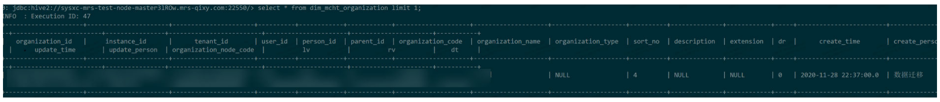
kinit *Component service user* (**kinit** is not required in a normal cluster.)

spark-sql

Step 2 Run the following command to set **spark.sql.hive.convertMetastoreOrc** to **false**:

set spark.sql.hive.convertMetastoreOrc=false;

Step 3 Query the Hive view again.



```
0: jdbc:hive2://sysxc-mrs-test-node-master31ROw.mrs-qixy.com:22550/> select * from dim_mcht_organization limit 1;
INFO  : Execution ID: 47
+------------------+-------------------+------------------+-----------+-------------+-----------+--------------------+--------------------+--------------------+----------+-------------+-------------+-----+------------------------+-----------
+-------------------+------------------+--------------------+------------+-------------+------------+
| organization_id  |    instance_id    |     tenant_id    | user_id  | person_id  | parent_id  | organization_code  | organization_name  | organization_type  | sort_no  | description  | extension  | dr  |      create_time       | create_perso
|   update_time     | update_person  | organization_node_code  |     lv     |      rv     |     dt     |
+------------------+-------------------+------------------+-----------+-------------+-----------+--------------------+--------------------+--------------------+----------+-------------+-------------+-----+------------------------+-----------
+-------------------+------------------+--------------------+------------+-------------+------------+
[redacted]  | NULL               | 4        | NULL        | NULL        | 0   | 2020-11-28 22:37:00.0  | 数据迁移
+------------------+-------------------+------------------+-----------+-------------+-----------+--------------------+--------------------+--------------------+----------+-------------+-------------+-----+------------------------+-----------
```

----End

15.17 Using Sqoop

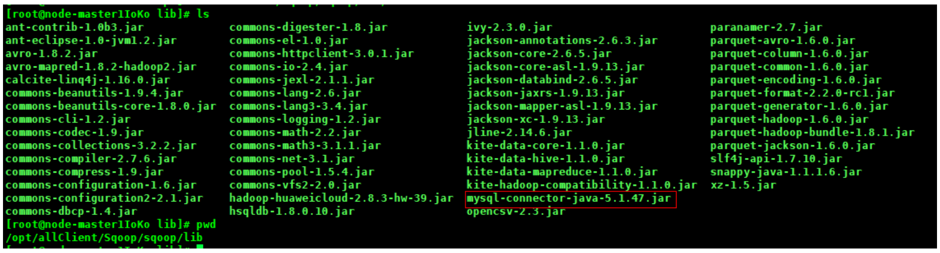
15.17.1 Connecting Sqoop to MySQL

Issue

The user does not know how to connect Sqoop to MySQL.

Procedure

Step 1 Install the client in the cluster and check whether the MySQL driver package exists in the **sqoop/lib** directory of the client.



```
[root@node-master1IoKo lib]# ls
ant-contrib-1.0b3.jar              commons-digester-1.8.jar             ivy-2.3.0.jar                         paranamer-2.7.jar
ant-eclipse-1.0-jvm1.2.jar         commons-el-1.0.jar                   jackson-annotations-2.6.3.jar         parquet-avro-1.6.0.jar
avro-1.8.2.jar                     commons-httpclient-3.0.1.jar         jackson-core-2.6.5.jar                parquet-column-1.6.0.jar
avro-mapred-1.8.2-hadoop2.jar      commons-io-2.4.jar                   jackson-core-asl-1.9.13.jar           parquet-common-1.6.0.jar
calcite-linq4j-1.16.0.jar          commons-jexl-2.1.1.jar               jackson-databind-2.6.5.jar            parquet-encoding-1.6.0.jar
commons-beanutils-1.9.4.jar        commons-lang-2.6.jar                 jackson-jaxrs-1.9.13.jar              parquet-format-2.2.0-rc1.jar
commons-beanutils-core-1.8.0.jar   commons-lang3-3.4.jar                jackson-mapper-asl-1.9.13.jar         parquet-generator-1.6.0.jar
commons-cli-1.2.jar                commons-logging-1.2.jar              jackson-xc-1.9.13.jar                 parquet-hadoop-1.6.0.jar
commons-codec-1.9.jar              commons-math-2.2.jar                 jline-2.14.6.jar                      parquet-hadoop-bundle-1.8.1.jar
commons-collections-3.2.2.jar      commons-math3-3.1.1.jar              kite-data-core-1.1.0.jar              parquet-jackson-1.6.0.jar
commons-compiler-2.7.6.jar         commons-net-3.1.jar                  kite-data-hive-1.1.0.jar              slf4j-api-1.7.10.jar
commons-compress-1.9.jar           commons-pool-1.5.4.jar               kite-data-mapreduce-1.1.0.jar         snappy-java-1.1.1.6.jar
commons-configuration-1.6.jar      commons-vfs2-2.0.jar                 kite-hadoop-compatibility-1.1.0.jar   xz-1.5.jar
commons-configuration2-2.1.jar     hadoop-huaweicloud-2.8.3-hw-39.jar   mysql-connector-java-5.1.47.jar
commons-dbcp-1.4.jar               hsqldb-1.8.0.10.jar                  opencsv-2.3.jar
[root@node-master1IoKo lib]# pwd
/opt/allClient/Sqoop/sqoop/lib
```

Step 2 Load environment variables in the client directory.

source bigdata_env

Step 3 Perform the Kerberos authentication.

If Kerberos authentication is not enabled for the cluster, skip this step. If it is enabled, run the following command to authenticate the current user:

kinit *MRS cluster user*

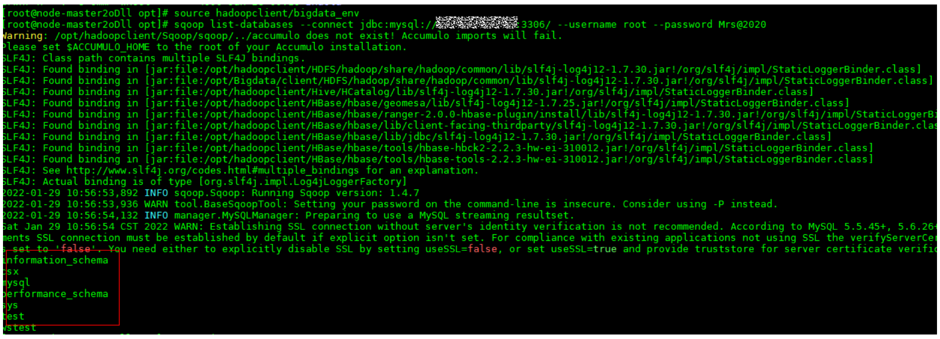
For example:

kinit admin

Step 4 Connect to the database.

sqoop list-databases --connect jdbc:mysql://IP:3306/ --username *Username* --password *Password*

An example is as follows.



```
[root@node-master2oDll opt]# source hadoopclient/bigdata_env
[root@node-master2oDll opt]# sqoop list-databases --connect jdbc:mysql://[redacted]:3306/ --username root --password Mrs@2020
Warning: /opt/hadoopclient/Sqoop/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/hadoopclient/HDFS/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.30.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/Bigdata/client/HDFS/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.30.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/hadoopclient/Hive/HCatalog/lib/slf4j-log4j12-1.7.30.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/hadoopclient/HBase/hbase/geomesa/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/hadoopclient/HBase/hbase/ranger-2.0.0-hbase-plugin/install/lib/slf4j-log4j12-1.7.30.jar!/org/slf4j/impl/StaticLoggerBi
SLF4J: Found binding in [jar:file:/opt/hadoopclient/HBase/hbase/lib/client-facing-thirdparty/slf4j-log4j12-1.7.30.jar!/org/slf4j/impl/StaticLoggerBinder.clas
SLF4J: Found binding in [jar:file:/opt/hadoopclient/HBase/hbase/lib/jdbc/slf4j-log4j12-1.7.30.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/hadoopclient/HBase/hbase/tools/hbase-hbck2-2.2.3-hw-ei-310012.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/hadoopclient/HBase/hbase/tools/hbase-tools-2.2.3-hw-ei-310012.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2022-01-29 10:56:53,892 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2022-01-29 10:56:53,936 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2022-01-29 10:56:54,132 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
Sat Jan 29 10:56:54 CST 2022 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+
ments SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCer
s set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verific
information_schema
csx
mysql
performance_schema
sys
test
wstest
```

The command output shows that Sqoop is successfully connected to the MySQL database.

----End

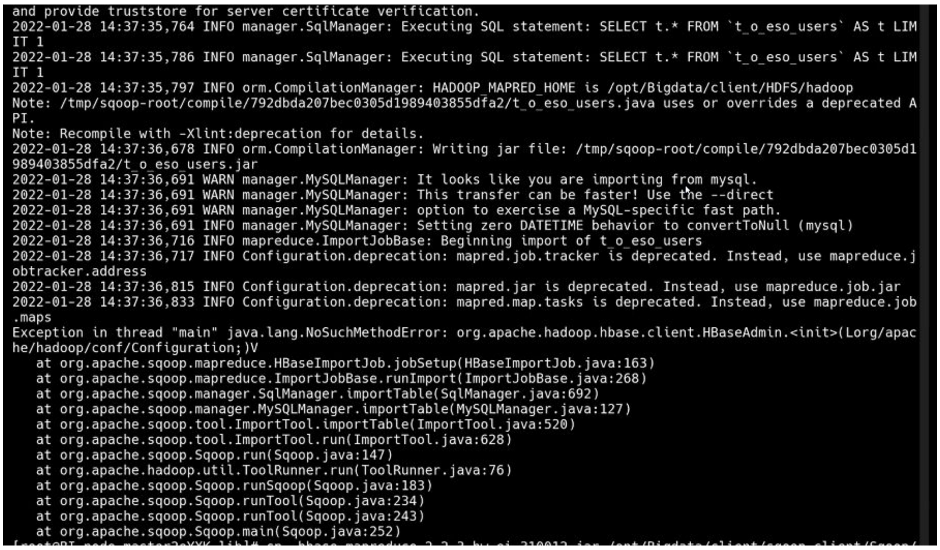
15.17.2 Failed to Find the HBaseAdmin.<init> Method When Sqoop Reads Data from the MySQL Database to HBase

Issue

If the Sqoop client (version 1.4.7) of MRS is used to extract data from a specified table in the MySQL database to a table in HBase 2.2.3, the following exception is reported:

```
Trying to load data into HBASE through Sqoop getting below error.
Exception in thread "main" java.lang.NoSuchMethodError:
org.apache.hadoop.hbase.client.HBaseAdmin.<init>(Lorg/apache/hadoop/conf/Configuration;)V
```

The following figure shows the complete exception information.



```
and provide truststore for server certificate verification.
2022-01-28 14:37:35,764 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `t_o_eso_users` AS t LIM
IT 1
2022-01-28 14:37:35,786 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `t_o_eso_users` AS t LIM
IT 1
2022-01-28 14:37:35,797 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /opt/Bigdata/client/HDFS/hadoop
Note: /tmp/sqoop-root/compile/792dbda207bec0305d1989403855dfa2/t_o_eso_users.java uses or overrides a deprecated A
PI.
Note: Recompile with -Xlint:deprecation for details.
2022-01-28 14:37:36,678 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/792dbda207bec0305d1
989403855dfa2/t_o_eso_users.jar
2022-01-28 14:37:36,691 WARN manager.MySQLManager: It looks like you are importing from mysql.
2022-01-28 14:37:36,691 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
2022-01-28 14:37:36,691 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
2022-01-28 14:37:36,691 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
2022-01-28 14:37:36,716 INFO mapreduce.ImportJobBase: Beginning import of t_o_eso_users
2022-01-28 14:37:36,717 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.j
obtracker.address
2022-01-28 14:37:36,815 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2022-01-28 14:37:36,833 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job
.maps
Exception in thread "main" java.lang.NoSuchMethodError: org.apache.hadoop.hbase.client.HBaseAdmin.<init>(Lorg/apac
he/hadoop/conf/Configuration;)V
    at org.apache.sqoop.mapreduce.HBaseImportJob.jobSetup(HBaseImportJob.java:163)
    at org.apache.sqoop.mapreduce.ImportJobBase.runImport(ImportJobBase.java:268)
    at org.apache.sqoop.manager.SqlManager.importTable(SqlManager.java:692)
    at org.apache.sqoop.manager.MySQLManager.importTable(MySQLManager.java:127)
    at org.apache.sqoop.tool.ImportTool.importTable(ImportTool.java:520)
    at org.apache.sqoop.tool.ImportTool.run(ImportTool.java:628)
    at org.apache.sqoop.Sqoop.run(Sqoop.java:147)
    at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)
    at org.apache.sqoop.Sqoop.runSqoop(Sqoop.java:183)
    at org.apache.sqoop.Sqoop.runTool(Sqoop.java:234)
    at org.apache.sqoop.Sqoop.runTool(Sqoop.java:243)
    at org.apache.sqoop.Sqoop.main(Sqoop.java:252)
```

The following is an example of running the Sqoop command to extract data:

```
sqoop import \
--connect jdbc:mysql://mysqlServer address:Port number/database1 \
--username admin \
--password xxx \
--table table1 \
--hbase-table table2 \
--column-family info \
--hbase-row-key id \
--hbase-create-table --m 1
```

Procedure

After the Sqoop client is installed, the JAR packages on which HBase depends are not imported. You need to manually import the JAR packages on which HBase of an earlier version depends.

Step 1 Check whether the Sqoop and HBase clients are in the same path.

- If yes, go to **Step 2**.
- If no, delete the original Sqoop and HBase client files, download the complete clients from FusionInsight Manager, and install them in the same path. Then go to **Step 2**.

Step 2 Log in to the node where the Sqoop client is installed as user **root**.

Step 3 Download JAR packages of HBase 1.6.0 and upload them to the **lib** directory on the Sqoop client:

Step 4 After the packages are uploaded, run the following command to change the permission on the packages to **755**:

chmod 755 *Package name*

Step 5 Run the following command in the client directory to refresh the Sqoop client:

source bigdata_env

Run the target Sqoop command again.

----End

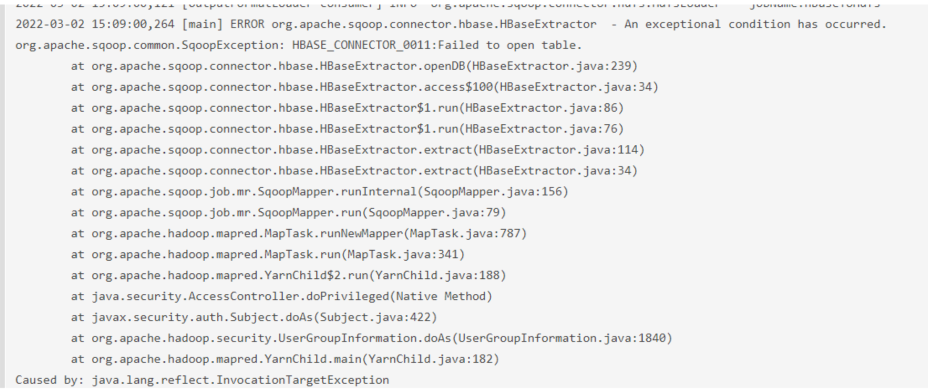
15.17.3 Failed to Export HBase Data to HDFS Through Hue's Sqoop Task

This section applies only to MRS 1.9.2 clusters.

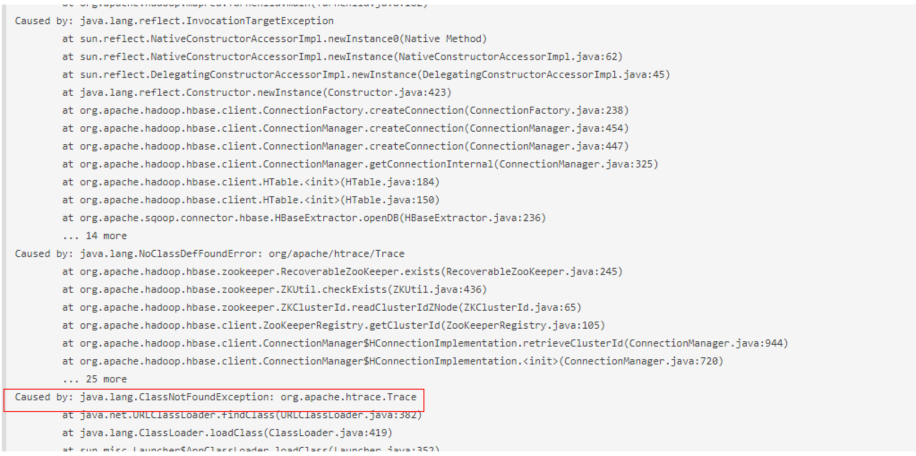
Issue

An error is reported when a Sqoop operation is performed on Hue to export data from HBase to HDFS.

Caused by: java.lang.ClassNotFoundException: org.apache.htrace.Trace



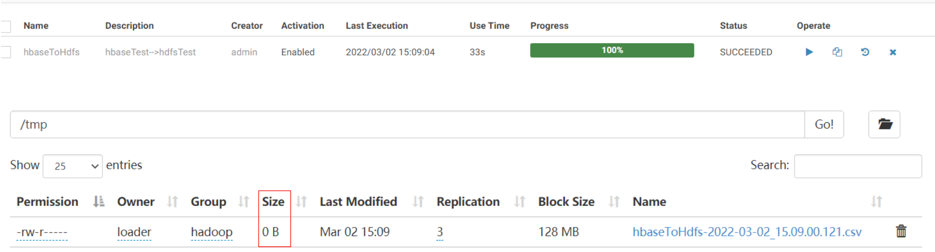
```
2022-03-02 15:09:00,264 [main] ERROR org.apache.sqoop.connector.hbase.HBaseExtractor  - An exceptional condition has occurred.
org.apache.sqoop.common.SqoopException: HBASE_CONNECTOR_0011:Failed to open table.
        at org.apache.sqoop.connector.hbase.HBaseExtractor.openDB(HBaseExtractor.java:239)
        at org.apache.sqoop.connector.hbase.HBaseExtractor.access$100(HBaseExtractor.java:34)
        at org.apache.sqoop.connector.hbase.HBaseExtractor$1.run(HBaseExtractor.java:86)
        at org.apache.sqoop.connector.hbase.HBaseExtractor$1.run(HBaseExtractor.java:76)
        at org.apache.sqoop.connector.hbase.HBaseExtractor.extract(HBaseExtractor.java:114)
        at org.apache.sqoop.connector.hbase.HBaseExtractor.extract(HBaseExtractor.java:34)
        at org.apache.sqoop.job.mr.SqoopMapper.runInternal(SqoopMapper.java:156)
        at org.apache.sqoop.job.mr.SqoopMapper.run(SqoopMapper.java:79)
        at org.apache.hadoop.mapred.MapTask.runNewMapper(MapTask.java:787)
        at org.apache.hadoop.mapred.MapTask.run(MapTask.java:341)
        at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:188)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1840)
        at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:182)
Caused by: java.lang.reflect.InvocationTargetException
```



```
Caused by: java.lang.reflect.InvocationTargetException
        at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native Method)
        at sun.reflect.NativeConstructorAccessorImpl.newInstance(NativeConstructorAccessorImpl.java:62)
        at sun.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:45)
        at java.lang.reflect.Constructor.newInstance(Constructor.java:423)
        at org.apache.hadoop.hbase.client.ConnectionFactory.createConnection(ConnectionFactory.java:238)
        at org.apache.hadoop.hbase.client.ConnectionManager.createConnection(ConnectionManager.java:454)
        at org.apache.hadoop.hbase.client.ConnectionManager.createConnection(ConnectionManager.java:447)
        at org.apache.hadoop.hbase.client.ConnectionManager.getConnectionInternal(ConnectionManager.java:325)
        at org.apache.hadoop.hbase.client.HTable.<init>(HTable.java:184)
        at org.apache.hadoop.hbase.client.HTable.<init>(HTable.java:150)
        at org.apache.sqoop.connector.hbase.HBaseExtractor.openDB(HBaseExtractor.java:236)
        ... 14 more
Caused by: java.lang.NoClassDefFoundError: org/apache/htrace/Trace
        at org.apache.hadoop.hbase.zookeeper.RecoverableZooKeeper.exists(RecoverableZooKeeper.java:245)
        at org.apache.hadoop.hbase.zookeeper.ZKUtil.checkExists(ZKUtil.java:436)
        at org.apache.hadoop.hbase.zookeeper.ZKClusterId.readClusterIdZNode(ZKClusterId.java:65)
        at org.apache.hadoop.hbase.client.ZooKeeperRegistry.getClusterId(ZooKeeperRegistry.java:105)
        at org.apache.hadoop.hbase.client.ConnectionManager$HConnectionImplementation.retrieveClusterId(ConnectionManager.java:944)
        at org.apache.hadoop.hbase.client.ConnectionManager$HConnectionImplementation.<init>(ConnectionManager.java:720)
        ... 25 more
Caused by: java.lang.ClassNotFoundException: org.apache.htrace.Trace
        at java.net.URLClassLoader.findClass(URLClassLoader.java:382)
        at java.lang.ClassLoader.loadClass(ClassLoader.java:419)
```

Symptom

The Sqoop task is executed successfully, but the CSV file in HDFS is empty.



	Name	Description	Creator	Activation	Last Execution	Use Time	Progress	Status	Operate
□	hbaseToHdfs	hbaseTest-->hdfsTest	admin	Enabled	2022/03/02 15:09:04	33s	100%	SUCCEEDED	

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r-----	loader	hadoop	0 B	Mar 02 15:09	3	128 MB	hbaseToHdfs-2022-03-02_15.09.00.121.csv

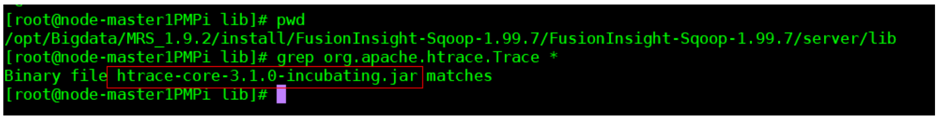
Cause Analysis

The JAR package conflicts or is missing.

Procedure

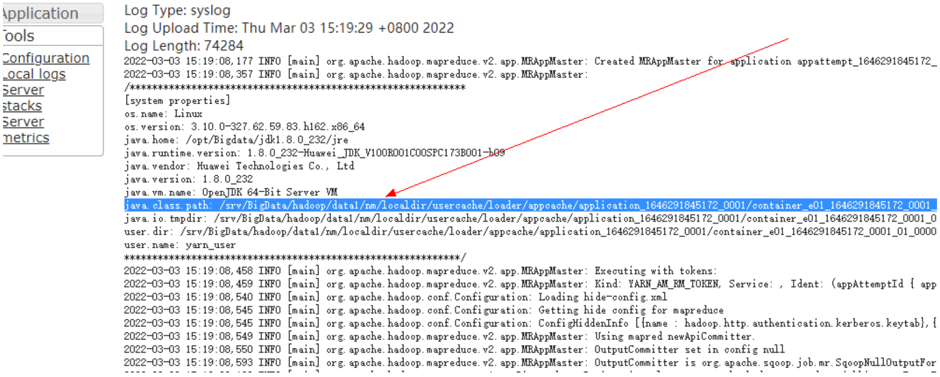
Step 1 Use the **grep** command in the **lib** directory of Sqoop.

1. Go to the **lib** directory of Sqoop and run the **grep** command.



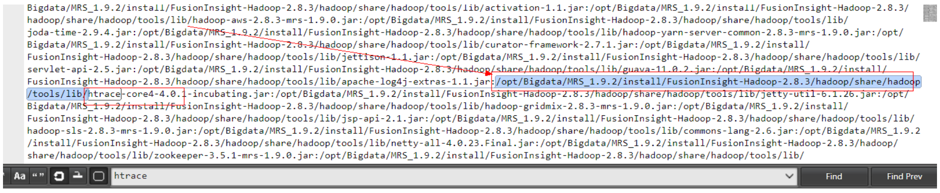
```
[root@node-master1PMPi lib]# pwd
/opt/Bigdata/MRS_1.9.2/install/FusionInsight-Sqoop-1.99.7/FusionInsight-Sqoop-1.99.7/server/lib
[root@node-master1PMPi lib]# grep org.apache.htrace.Trace *
Binary file htrace-core-3.1.0-incubating.jar matches
[root@node-master1PMPi lib]#
```

2. Go to the native Yarn page and view the error information about the running task.



```
Log Type: syslog
Log Upload Time: Thu Mar 03 15:19:29 +0800 2022
Log Length: 74284
2022-03-03 15:19:08,177 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Created MRAppMaster for application appattempt_1646291845172_
2022-03-03 15:19:08,357 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster:
/************************************************************
[system properties]
os.name: Linux
os.version: 3.10.0-327.62.59.83.h162.x86_64
java.home: /opt/Bigdata/jdk1.8.0_232/jre
java.runtime.version: 1.8.0_232-Huawei_JDK_V100R001C00SPC173B001-b09
java.vendor: Huawei Technologies Co., Ltd
java.version: 1.8.0_232
java.vm.name: OpenJDK 64-Bit Server VM
java.class.path: /srv/BigData/hadoop/data1/nm/localdir/usercache/loader/appcache/application_1646291845172_0001/container_e01_1646291845172_0001_
java.io.tmpdir: /srv/BigData/hadoop/data1/nm/localdir/usercache/loader/appcache/application_1646291845172_0001/container_e01_1646291845172_0001_0
user.dir: /srv/BigData/hadoop/data1/nm/localdir/usercache/loader/appcache/application_1646291845172_0001/container_e01_1646291845172_0001_01_0000
user.name: yarn_user
************************************************************/
2022-03-03 15:19:08,458 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Executing with tokens:
2022-03-03 15:19:08,459 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Kind: YARN_AM_RM_TOKEN, Service: , Ident: (appAttemptId { app
2022-03-03 15:19:08,540 INFO [main] org.apache.hadoop.conf.Configuration: Loading hide-config.xml
2022-03-03 15:19:08,545 INFO [main] org.apache.hadoop.conf.Configuration: Getting hide config for mapreduce
2022-03-03 15:19:08,545 INFO [main] org.apache.hadoop.conf.Configuration: ConfigHiddenInfo [{name : hadoop.http.authentication.kerberos.keytab}, {
2022-03-03 15:19:08,549 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: Using mapred newApiCommitter.
2022-03-03 15:19:08,550 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: OutputCommitter set in config null
2022-03-03 15:19:08,593 INFO [main] org.apache.hadoop.mapreduce.v2.app.MRAppMaster: OutputCommitter is org.apache.sqoop.job.mr.SqoopNullOutputFor
```

3. Copy **java.class.path** and search for **htrace-core**.



```
Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/activation-1.1.jar:/opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/
hadoop/share/hadoop/tools/lib/hadoop-aws-2.8.3-mrs-1.9.0.jar:/opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/
joda-time-2.9.4.jar:/opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/hadoop-yarn-server-common-2.8.3-mrs-1.9.0.jar:/opt/
Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/curator-framework-2.7.1.jar:/opt/Bigdata/MRS_1.9.2/install/
FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/jettison-1.1.jar:/opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/
servlet-api-2.5.jar:/opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/guava-11.0.2.jar:/opt/Bigdata/MRS_1.9.2/install/
FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/apache-log4j-extras-1.1.jar:/opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop
/tools/lib/htrace-core4-4.0.1-incubating.jar:/opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/jetty-util-6.1.26.jar:/opt/
Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/hadoop-gridmix-2.8.3-mrs-1.9.0.jar:/opt/Bigdata/MRS_1.9.2/install/
FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/jsp-api-2.1.jar:/opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/
hadoop-sls-2.8.3-mrs-1.9.0.jar:/opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/commons-lang-2.6.jar:/opt/Bigdata/MRS_1.9.2
/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/netty-all-4.0.23.Final.jar:/opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/
share/hadoop/tools/lib/zookeeper-3.5.1-mrs-1.9.0.jar:/opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/tools/lib/
```

4. Copy the JAR package to the following directory:

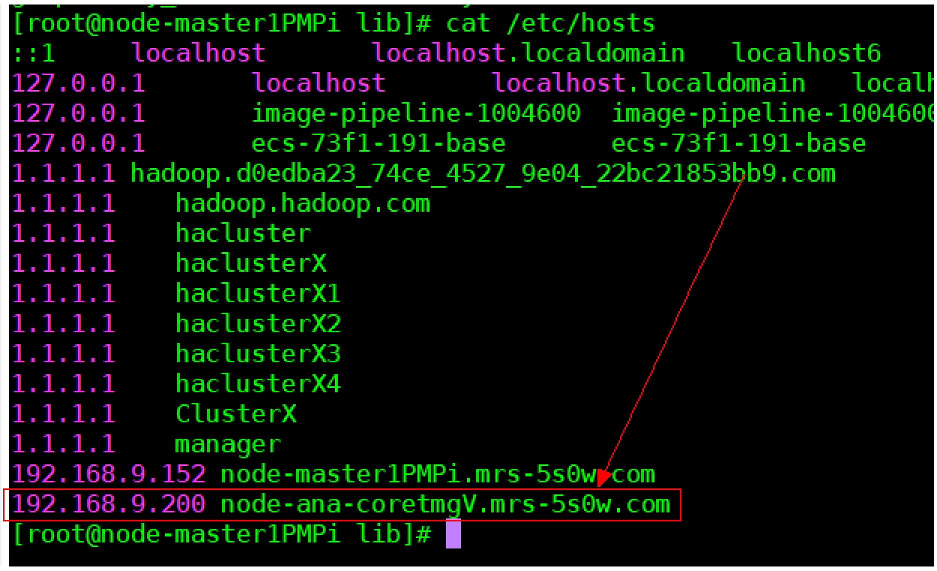
cp /opt/Bigdata/MRS_1.9.2/install/FusionInsight-Sqoop-1.99.7/FusionInsight-Sqoop-1.99.7/server/lib/htrace-core-3.1.0-incubating.jar /opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/common/lib/

5. Change permissions.

chmod 777 *htrace-core-3.1.0-incubating.jar* (the copied JAR package)

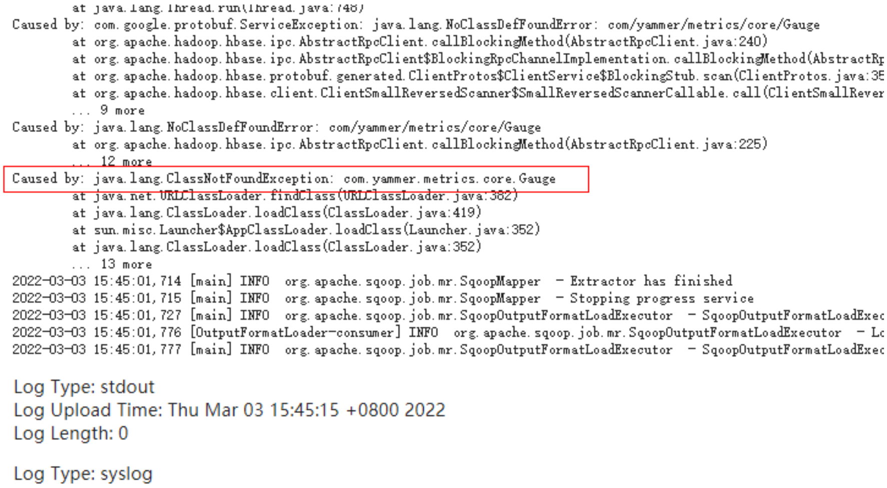
chown omm:ficommon *htrace-core-3.1.0-incubating.jar* (the copied JAR package)

6. View the **hosts** file and perform the same operations to copy the JAR package for all other nodes.



```
[root@node-master1PMPi lib]# cat /etc/hosts
::1     localhost       localhost.localdomain   localhost6
127.0.0.1       localhost       localhost.localdomain   localh
127.0.0.1       image-pipeline-1004600  image-pipeline-100460
127.0.0.1       ecs-73f1-191-base       ecs-73f1-191-base
1.1.1.1 hadoop.d0edba23_74ce_4527_9e04_22bc21853bb9.com
1.1.1.1    hadoop.hadoop.com
1.1.1.1    hacluster
1.1.1.1    haclusterX
1.1.1.1    haclusterX1
1.1.1.1    haclusterX2
1.1.1.1    haclusterX3
1.1.1.1    haclusterX4
1.1.1.1    ClusterX
1.1.1.1    manager
192.168.9.152 node-master1PMPi.mrs-5s0w.com
192.168.9.200 node-ana-coretmgV.mrs-5s0w.com
[root@node-master1PMPi lib]#
```

7. Run the Sqoop task again. The following error information is displayed.



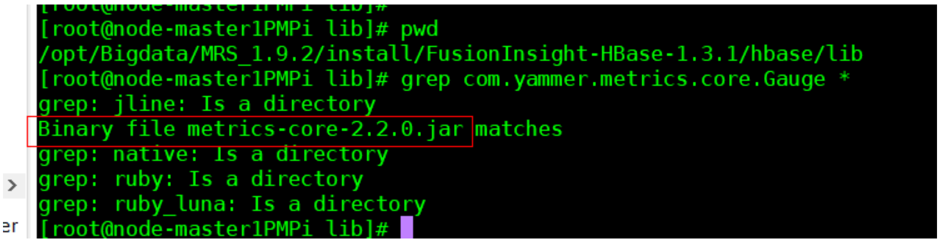
```
        at java.lang.Thread.run(Thread.java:748)
Caused by: com.google.protobuf.ServiceException: java.lang.NoClassDefFoundError: com/yammer/metrics/core/Gauge
        at org.apache.hadoop.hbase.ipc.AbstractRpcClient.callBlockingMethod(AbstractRpcClient.java:240)
        at org.apache.hadoop.hbase.ipc.AbstractRpcClient$BlockingRpcChannelImplementation.callBlockingMethod(AbstractRp
        at org.apache.hadoop.hbase.protobuf.generated.ClientProtos$ClientService$BlockingStub.scan(ClientProtos.java:35
        at org.apache.hadoop.hbase.client.ClientSmallReversedScanner$SmallReversedScannerCallable.call(ClientSmallRever
        ... 9 more
Caused by: java.lang.NoClassDefFoundError: com/yammer/metrics/core/Gauge
        at org.apache.hadoop.hbase.ipc.AbstractRpcClient.callBlockingMethod(AbstractRpcClient.java:225)
        ... 12 more
Caused by: java.lang.ClassNotFoundException: com.yammer.metrics.core.Gauge
        at java.net.URLClassLoader.findClass(URLClassLoader.java:382)
        at java.lang.ClassLoader.loadClass(ClassLoader.java:419)
        at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:352)
        at java.lang.ClassLoader.loadClass(ClassLoader.java:352)
        ... 13 more
2022-03-03 15:45:01,714 [main] INFO  org.apache.sqoop.job.mr.SqoopMapper  - Extractor has finished
2022-03-03 15:45:01,715 [main] INFO  org.apache.sqoop.job.mr.SqoopMapper  - Stopping progress service
2022-03-03 15:45:01,727 [main] INFO  org.apache.sqoop.job.mr.SqoopOutputFormatLoadExecutor  - SqoopOutputFormatLoadExec
2022-03-03 15:45:01,776 [OutputFormatLoader-consumer] INFO  org.apache.sqoop.job.mr.SqoopOutputFormatLoadExecutor  - Lc
2022-03-03 15:45:01,777 [main] INFO  org.apache.sqoop.job.mr.SqoopOutputFormatLoadExecutor  - SqoopOutputFormatLoadExec
```

Log Type: stdout
Log Upload Time: Thu Mar 03 15:45:15 +0800 2022
Log Length: 0

Log Type: syslog

Step 2 Use the **grep** command in the **lib** directory of HBase.

1. Go to the **lib** directory of HBase and run the **grep** command.



```
[root@node-master1PMPi lib]#
[root@node-master1PMPi lib]# pwd
/opt/Bigdata/MRS_1.9.2/install/FusionInsight-HBase-1.3.1/hbase/lib
[root@node-master1PMPi lib]# grep com.yammer.metrics.core.Gauge *
grep: jline: Is a directory
Binary file metrics-core-2.2.0.jar matches
grep: native: Is a directory
grep: ruby: Is a directory
grep: ruby_luna: Is a directory
[root@node-master1PMPi lib]#
```

2. Copy the JAR package.

   cp /opt/Bigdata/MRS_1.9.2/install/FusionInsight-HBase-1.3.1/hbase/lib/metrics-core-2.2.0.jar /opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/common/lib/

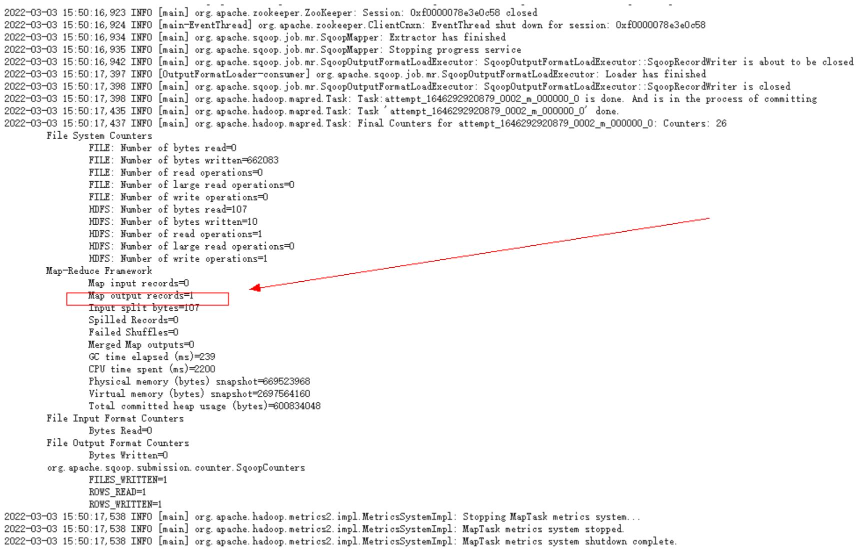
3. Change permissions.

   chmod 777 *metrics-core-2.2.0.jar* (the copied JAR package)

   chown omm:ficommon *metrics-core-2.2.0.jar* (the copied JAR package)

4. View the **hosts** file and perform the same operations to copy the JAR package for all other nodes.

5. Run the Sqoop task.



```
2022-03-03 15:50:16,923 INFO [main] org.apache.zookeeper.ZooKeeper: Session: 0xf0000078e3e0c58 closed
2022-03-03 15:50:16,924 INFO [main-EventThread] org.apache.zookeeper.ClientCnxn: EventThread shut down for session: 0xf0000078e3e0c58
2022-03-03 15:50:16,934 INFO [main] org.apache.sqoop.job.mr.SqoopMapper: Extractor has finished
2022-03-03 15:50:16,935 INFO [main] org.apache.sqoop.job.mr.SqoopMapper: Stopping progress service
2022-03-03 15:50:16,942 INFO [main] org.apache.sqoop.job.mr.SqoopOutputFormatLoadExecutor: SqoopOutputFormatLoadExecutor::SqoopRecordWriter is about to be closed
2022-03-03 15:50:17,397 INFO [OutputFormatLoader-consumer] org.apache.sqoop.job.mr.SqoopOutputFormatLoadExecutor: Loader has finished
2022-03-03 15:50:17,398 INFO [main] org.apache.sqoop.job.mr.SqoopOutputFormatLoadExecutor: SqoopOutputFormatLoadExecutor::SqoopRecordWriter is closed
2022-03-03 15:50:17,398 INFO [main] org.apache.hadoop.mapred.Task: Task:attempt_1646292920879_0002_m_000000_0 is done. And is in the process of commiting
2022-03-03 15:50:17,435 INFO [main] org.apache.hadoop.mapred.Task: Task 'attempt_1646292920879_0002_m_000000_0' done.
2022-03-03 15:50:17,437 INFO [main] org.apache.hadoop.mapred.Task: Final Counters for attempt_1646292920879_0002_m_000000_0: Counters: 26
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=662083
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=107
                HDFS: Number of bytes written=10
                HDFS: Number of read operations=1
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=1
        Map-Reduce Framework
                Map input records=0
                Map output records=1
                Input split bytes=107
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=239
                CPU time spent (ms)=2200
                Physical memory (bytes) snapshot=669523968
                Virtual memory (bytes) snapshot=2697564160
                Total committed heap usage (bytes)=600834048
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
        org.apache.sqoop.submission.counter.SqoopCounters
                FILES_WRITTEN=1
                ROWS_READ=1
                ROWS_WRITTEN=1
2022-03-03 15:50:17,538 INFO [main] org.apache.hadoop.metrics2.impl.MetricsSystemImpl: Stopping MapTask metrics system...
2022-03-03 15:50:17,538 INFO [main] org.apache.hadoop.metrics2.impl.MetricsSystemImpl: MapTask metrics system stopped.
2022-03-03 15:50:17,538 INFO [main] org.apache.hadoop.metrics2.impl.MetricsSystemImpl: MapTask metrics system shutdown complete.
```

----End

Conclusion

1. Copy **htrace-core-3.1.0-incubating.jar** in the **lib** directory of Sqoop and **metrics-core-2.2.0.jar** in the **lib** directory of HBase to **/opt/Bigdata/MRS_1.9.2/install/FusionInsight-Hadoop-2.8.3/hadoop/share/hadoop/common/lib/**.
2. Change the permissions for the JAR packages to **777** and **omm:ficommon**, respectively.
3. Perform the preceding operations on all nodes and run the Sqoop task again.

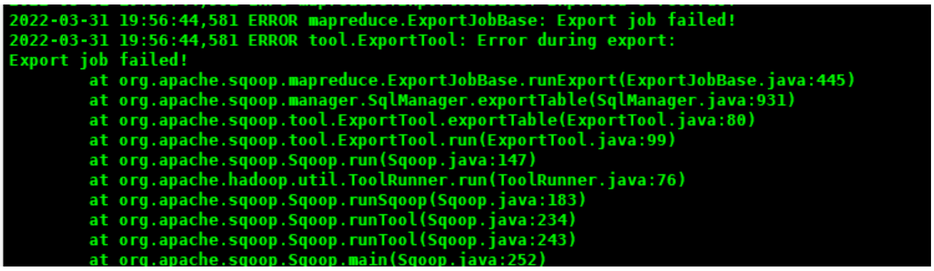
15.17.4 A Format Error Is Reported When Sqoop Is Used to Export Data from Hive to MySQL 8.0

This section applies only to MRS 3.1.0 clusters.

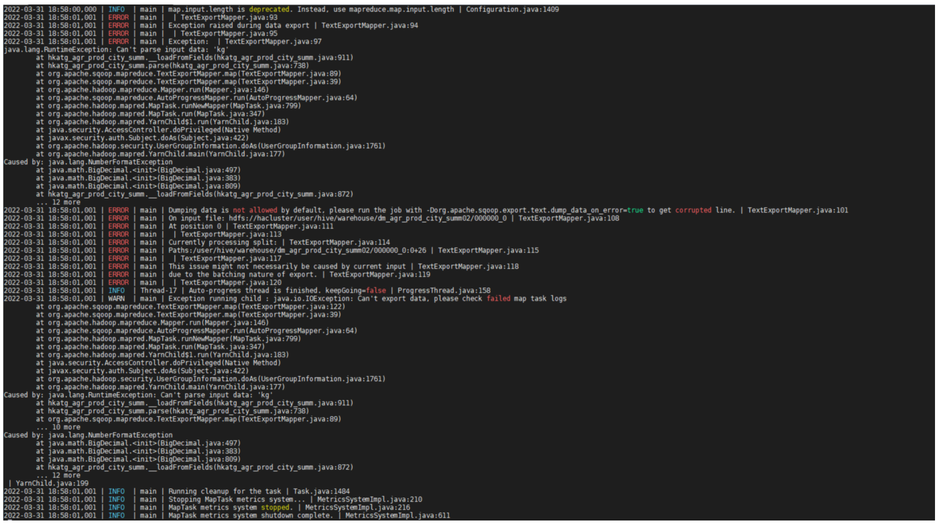
Issue

A format error is reported when a Sqoop task is performed to export data from Hive to MySQL 8.0 in an MRS 3.1.0 cluster.

Symptom



```
2022-03-31 19:56:44,581 ERROR mapreduce.ExportJobBase: Export job failed!
2022-03-31 19:56:44,581 ERROR tool.ExportTool: Error during export:
Export job failed!
        at org.apache.sqoop.mapreduce.ExportJobBase.runExport(ExportJobBase.java:445)
        at org.apache.sqoop.manager.SqlManager.exportTable(SqlManager.java:931)
        at org.apache.sqoop.tool.ExportTool.exportTable(ExportTool.java:80)
        at org.apache.sqoop.tool.ExportTool.run(ExportTool.java:99)
        at org.apache.sqoop.Sqoop.run(Sqoop.java:147)
        at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)
        at org.apache.sqoop.Sqoop.runSqoop(Sqoop.java:183)
        at org.apache.sqoop.Sqoop.runTool(Sqoop.java:234)
        at org.apache.sqoop.Sqoop.runTool(Sqoop.java:243)
        at org.apache.sqoop.Sqoop.main(Sqoop.java:252)
```



```
2022-03-31 18:58:00,000 | INFO  | main | map.input.length is deprecated. Instead, use mapreduce.map.input.length | Configuration.java:1409
2022-03-31 18:58:01,001 | ERROR | main |  | TextExportMapper.java:93
2022-03-31 18:58:01,001 | ERROR | main | Exception raised during data export | TextExportMapper.java:94
2022-03-31 18:58:01,001 | ERROR | main |  | TextExportMapper.java:95
2022-03-31 18:58:01,001 | ERROR | main | Exception:  | TextExportMapper.java:97
java.lang.RuntimeException: Can't parse input data: 'kg'
        at hkatg_agr_prod_city_summ.__loadFromFields(hkatg_agr_prod_city_summ.java:911)
        at hkatg_agr_prod_city_summ.parse(hkatg_agr_prod_city_summ.java:738)
        at org.apache.sqoop.mapreduce.TextExportMapper.map(TextExportMapper.java:89)
        at org.apache.sqoop.mapreduce.TextExportMapper.map(TextExportMapper.java:39)
        at org.apache.hadoop.mapreduce.Mapper.run(Mapper.java:146)
        at org.apache.sqoop.mapreduce.AutoProgressMapper.run(AutoProgressMapper.java:64)
        at org.apache.hadoop.mapred.MapTask.runNewMapper(MapTask.java:799)
        at org.apache.hadoop.mapred.MapTask.run(MapTask.java:347)
        at org.apache.hadoop.mapred.YarnChild$1.run(YarnChild.java:183)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1761)
        at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:177)
Caused by: java.lang.NumberFormatException
        at java.math.BigDecimal.<init>(BigDecimal.java:497)
        at java.math.BigDecimal.<init>(BigDecimal.java:383)
        at java.math.BigDecimal.<init>(BigDecimal.java:809)
        at hkatg_agr_prod_city_summ.__loadFromFields(hkatg_agr_prod_city_summ.java:872)
        ... 12 more
2022-03-31 18:58:01,001 | ERROR | main | Dumping data is not allowed by default, please run the job with -Dorg.apache.sqoop.export.text.dump_data_on_error=true to get corrupted line. | TextExportMapper.java:101
2022-03-31 18:58:01,001 | ERROR | main | On input file: hdfs://hacluster/user/hive/warehouse/dm_agr_prod_city_summ02/000000_0 | TextExportMapper.java:108
2022-03-31 18:58:01,001 | ERROR | main | At position 0 | TextExportMapper.java:111
2022-03-31 18:58:01,001 | ERROR | main |  | TextExportMapper.java:113
2022-03-31 18:58:01,001 | ERROR | main | Currently processing split: | TextExportMapper.java:114
2022-03-31 18:58:01,001 | ERROR | main | Paths:/user/hive/warehouse/dm_agr_prod_city_summ02/000000_0:0+26 | TextExportMapper.java:115
2022-03-31 18:58:01,001 | ERROR | main |  | TextExportMapper.java:117
2022-03-31 18:58:01,001 | ERROR | main | This issue might not necessarily be caused by current input | TextExportMapper.java:118
2022-03-31 18:58:01,001 | ERROR | main | due to the batching nature of export. | TextExportMapper.java:119
2022-03-31 18:58:01,001 | ERROR | main |  | TextExportMapper.java:120
2022-03-31 18:58:01,001 | INFO  | Thread-17 | Auto-progress thread is finished. keepGoing=false | ProgressThread.java:158
2022-03-31 18:58:01,001 | WARN  | main | Exception running child : java.io.IOException: Can't export data, please check failed map task logs
        at org.apache.sqoop.mapreduce.TextExportMapper.map(TextExportMapper.java:122)
        at org.apache.sqoop.mapreduce.TextExportMapper.map(TextExportMapper.java:39)
        at org.apache.hadoop.mapreduce.Mapper.run(Mapper.java:146)
        at org.apache.sqoop.mapreduce.AutoProgressMapper.run(AutoProgressMapper.java:64)
        at org.apache.hadoop.mapred.MapTask.runNewMapper(MapTask.java:799)
        at org.apache.hadoop.mapred.MapTask.run(MapTask.java:347)
        at org.apache.hadoop.mapred.YarnChild$1.run(YarnChild.java:183)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1761)
        at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:177)
Caused by: java.lang.RuntimeException: Can't parse input data: 'kg'
        at hkatg_agr_prod_city_summ.__loadFromFields(hkatg_agr_prod_city_summ.java:911)
        at hkatg_agr_prod_city_summ.parse(hkatg_agr_prod_city_summ.java:738)
        at org.apache.sqoop.mapreduce.TextExportMapper.map(TextExportMapper.java:89)
        ... 10 more
Caused by: java.lang.NumberFormatException
        at java.math.BigDecimal.<init>(BigDecimal.java:497)
        at java.math.BigDecimal.<init>(BigDecimal.java:383)
        at java.math.BigDecimal.<init>(BigDecimal.java:809)
        at hkatg_agr_prod_city_summ.__loadFromFields(hkatg_agr_prod_city_summ.java:872)
        ... 12 more
 | YarnChild.java:199
2022-03-31 18:58:01,001 | INFO  | main | Running cleanup for the task | Task.java:1484
2022-03-31 18:58:01,001 | INFO  | main | Stopping MapTask metrics system... | MetricsSystemImpl.java:210
2022-03-31 18:58:01,001 | INFO  | main | MapTask metrics system stopped. | MetricsSystemImpl.java:216
2022-03-31 18:58:01,001 | INFO  | main | MapTask metrics system shutdown complete. | MetricsSystemImpl.java:611
```

Cause Analysis

The log shows that the format is incorrect.

Procedure

Check that the formats of delimiters and table fields are all correct. Add the **--columns** parameter to the Sqoop statement to make data formats consistent between source and target tables.

sqoop export --connect jdbc:mysql://IP address:Port number/Database name --username *Username* --password *Password* --table *Table name* --columns *Column fields (multiple columns are separated by commas)* -export-dir *Export address* --fields-terminated-by *Delimiter* --input-null-string '\\N' --input-null-non-string '\\N' -m 1

Example:

sqoop export --connect jdbc:mysql://172.16.0.6:3306/lidengpeng --username root --password Mrs@2021 --table hkatg_agr_prod_city_summ --columns year,city_name,city_code,prod_code,prod_name,prod_type,sown_area,area_unit,yield_wegt,yield_unit,total_wegt,total_wegt_unit,data_sorc_code,etl_time -export-dir hdfs://hacluster/user/hive/warehouse/dm_agr_prod_city_summ02 --fields-terminated-by ',' --input-null-string '\\N' --input-null-non-string '\\N' -m 1

15.17.5 An Error Is Reported When sqoop import Is Executed to Import PostgreSQL Data to Hive

Background

The **sqoop import** command is executed to extract data from open-source PostgreSQL to MRS HDFS or Hive.

Issue

The **sqoop** command can be executed to query the PostgreSQL database table, but an error is reported when the **sqoop import** command is executed to import data.

The authentication type 5 is not supported. Check that you have configured the pg_hba.conf file to include the client's IP address or subnet.

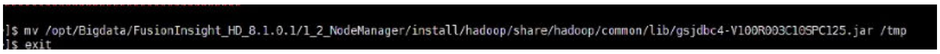
Cause Analysis

1. MD5 authentication for connecting to PostgreSQL fails. A whitelist needs to be configured in the **pg_hba.cnf** file.
2. When the **sqoop import** command is executed, a MapReduce job is started. The PostgreSQL driver package **gsjdbc4-*.jar** exists in the MRS Hadoop installation directory **/opt/Bigdata/FusionInsight_HD_*/1_*_DataNode/install/hadoop/share/hadoop/common/lib**, which is incompatible with the open-source PostgreSQL service. As a result, an error is reported.

Procedure

1. Configure a whitelist in the **pg_hba.cnf** file.
2. Delete the **gsjdbc4-*.jar** packages from all core nodes, and add the PostgreSQL JAR package to **sqoop/lib**.

   mv /opt/Bigdata/FusionInsight_HD_*/1_*_DataNode/install/hadoop/share/hadoop/common/lib/gsjdbc4-*.jar /tmp

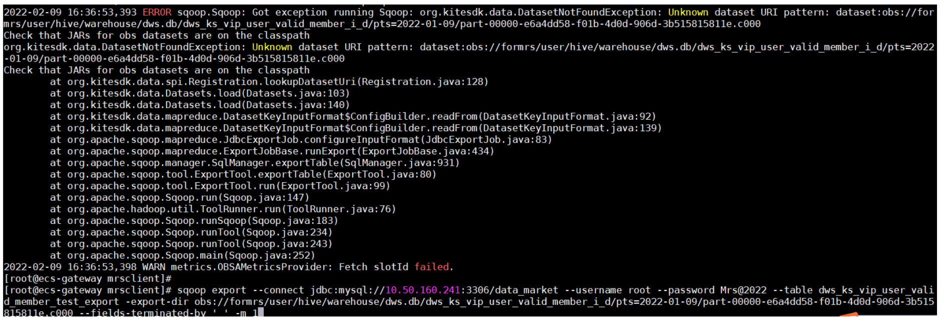
```
]$ mv /opt/Bigdata/FusionInsight_HD_8.1.0.1/1_2_NodeManager/install/hadoop/share/hadoop/common/lib/gsjdbc4-V100R003C10SPC125.jar /tmp
]$ exit
```

15.17.6 Sqoop Failed to Read Data from MySQL and Write Parquet Files to OBS

Issue

An error is reported when Sqoop reads MySQL data and writes the data to OBS in Parquet format. However, the data can be successfully written to OBS if the Parquet format is not specified.

Symptom



```
2022-02-09 16:36:53,393 ERROR sqoop.Sqoop: Got exception running Sqoop: org.kitesdk.data.DatasetNotFoundException: Unknown dataset URI pattern: dataset:obs://for
mrs/user/hive/warehouse/dws.db/dws_ks_vip_user_valid_member_i_d/pts=2022-01-09/part-00000-e6a4dd58-f01b-4d0d-906d-3b515815811e.c000
Check that JARs for obs datasets are on the classpath
org.kitesdk.data.DatasetNotFoundException: Unknown dataset URI pattern: dataset:obs://formrs/user/hive/warehouse/dws.db/dws_ks_vip_user_valid_member_i_d/pts=2022
-01-09/part-00000-e6a4dd58-f01b-4d0d-906d-3b515815811e.c000
Check that JARs for obs datasets are on the classpath
        at org.kitesdk.data.spi.Registration.lookupDatasetUri(Registration.java:128)
        at org.kitesdk.data.Datasets.load(Datasets.java:103)
        at org.kitesdk.data.Datasets.load(Datasets.java:140)
        at org.kitesdk.data.mapreduce.DatasetKeyInputFormat$ConfigBuilder.readFrom(DatasetKeyInputFormat.java:92)
        at org.kitesdk.data.mapreduce.DatasetKeyInputFormat$ConfigBuilder.readFrom(DatasetKeyInputFormat.java:139)
        at org.apache.sqoop.mapreduce.JdbcExportJob.configureInputFormat(JdbcExportJob.java:83)
        at org.apache.sqoop.mapreduce.ExportJobBase.runExport(ExportJobBase.java:434)
        at org.apache.sqoop.manager.SqlManager.exportTable(SqlManager.java:931)
        at org.apache.sqoop.tool.ExportTool.exportTable(ExportTool.java:80)
        at org.apache.sqoop.tool.ExportTool.run(ExportTool.java:99)
        at org.apache.sqoop.Sqoop.run(Sqoop.java:147)
        at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)
        at org.apache.sqoop.Sqoop.runSqoop(Sqoop.java:183)
        at org.apache.sqoop.Sqoop.runTool(Sqoop.java:234)
        at org.apache.sqoop.Sqoop.runTool(Sqoop.java:243)
        at org.apache.sqoop.Sqoop.main(Sqoop.java:252)
2022-02-09 16:36:53,398 WARN metrics.OBSAMetricsProvider: Fetch slotId failed.
[root@ecs-gateway mrsclient]#
[root@ecs-gateway mrsclient]# sqoop export --connect jdbc:mysql://10.50.160.241:3306/data_market --username root --password Mrs@2022 --table dws_ks_vip_user_vali
d_member_test_export -export-dir obs://formrs/user/hive/warehouse/dws.db/dws_ks_vip_user_valid_member_i_d/pts=2022-01-09/part-00000-e6a4dd58-f01b-4d0d-906d-3b515
815811e.c000 --fields-terminated-by ' ' -m 1
```

Cause Analysis

Parquet does not support Hive 3. Data can be written using HCatalog.

Procedure

Use HCatalog to write data: Specify the Hive database and table in parameters and modify the SQL statement in the script.

Details are as follows:

Original script:

```
sqoop import --connect 'jdbc:mysql://10.160.5.65/xxx_pos_online_00?zeroDateTimeBehavior=convertToNull' --username root --password Mrs@2022

--split-by id

--num-mappers 2

--query 'select * from pos_remark where 1=1 and $CONDITIONS'

--target-dir obs://za-test/dev/xxx_pos_online_00/pos_remark

--delete-target-dir

--null-string '\\N'

--null-non-string '\\N'

--as-parquetfile
```

Modified script:

```
sqoop import --connect 'jdbc:mysql://10.160.5.65/xxx_pos_online_00?zeroDateTimeBehavior=convertToNull' --username root --password Mrs@2022

--split-by id

--num-mappers 2

--query 'select id,pos_case_id,pos_transaction_id,remark,update_time,update_user,is_deleted,creator,modifier,gmt_created,gmt_modified,update_user_id,tenant_code from pos_remark where 1=1 and $CONDITIONS'

--hcatalog-database xxx_dev

--hcatalog-table ods_pos_remark
```

15.18 Using Storm

15.18.1 Invalid Hyperlink of Events on the Storm UI

Issue

The hyperlink of events on the Storm UI is invalid.

Symptom

After submitting a topology, a user cannot view topology data processing logs and the events hyperlink is invalid.

Cause Analysis

The function of viewing topology data processing logs is disabled by default when a topology is submitted in an MRS cluster.

Procedure

Step 1 Go to the service page.

- For versions earlier than MRS 2.0.1: Log in to MRS Manager and choose **Services**.
- For MRS 2.0.1 or later: Click the cluster name on the MRS console and choose **Components**.

  NOTE

  If the **Components** tab is unavailable, complete IAM user synchronization first. (On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.)
- For MRS 3.*x* or later: Log in to FusionInsight Manager and choose **Cluster**. Click the name of the desired cluster and choose **Services**.

Step 2 Log in to the Storm web UI.

- For MRS 2.x and earlier versions: Choose **Storm**. On the **Storm WebUI** page, click any UI link to open the Storm web UI.

  NOTE

  When accessing the Storm web UI for the first time, you must add the address to the trusted site list.
- For MRS 3.x or later: Choose **Storm** > **Overview**. On the **Storm WebUI** in the **Basic Information** area, click any UI link to open the Storm web UI.

Step 3 In the **Topology Summary** area, click the desired topology to view details.

Step 4 In the **Topology actions** area, click **Kill** to delete the submitted Storm topology.

Step 5 Submit the Storm topology again and enable the function of viewing topology data processing logs. Add the **topology.eventlogger.executors** parameter and set it to a positive integer when submitting the Storm topology. Example:

storm jar *Path of the topology package Class name of the topology Main method Topology name* -c topology.eventlogger.executors=X

Step 6 In the **Topology Summary** area on the Storm UI, click the desired topology to view details.

Step 7 In the **Topology actions** area, click **Debug**, specify the data sampling percentage, and click **OK**.

Step 8 Click the **Spouts** or **Bolts** task name of the topology. In **Component summary**, click **events** to view data processing logs.

NOTE

To enable the function of viewing topology data processing logs of the specified **Spouts** or **Bolts** task, click the **Spouts** or **Bolts** task name of the topology, click **Debug** in the **Topology actions** area, and enter the data sampling percentage.

----End

15.18.2 Failed to Submit a Topology

Symptom

An MRS streaming cluster is installed, and ZooKeeper, Storm, as well as Kafka are installed in the cluster.

A topology fails to be submitted by running commands on the client.

Possible Causes

- The Storm service is abnormal.
- The client user is not authenticated or the authentication has expired.
- The **storm.yaml** file in the submitted topology conflicts with that on the server.

Cause Analysis

A user fails to submit the topology. The possible cause is that the client or Storm is faulty.

1. Check the Storm status.

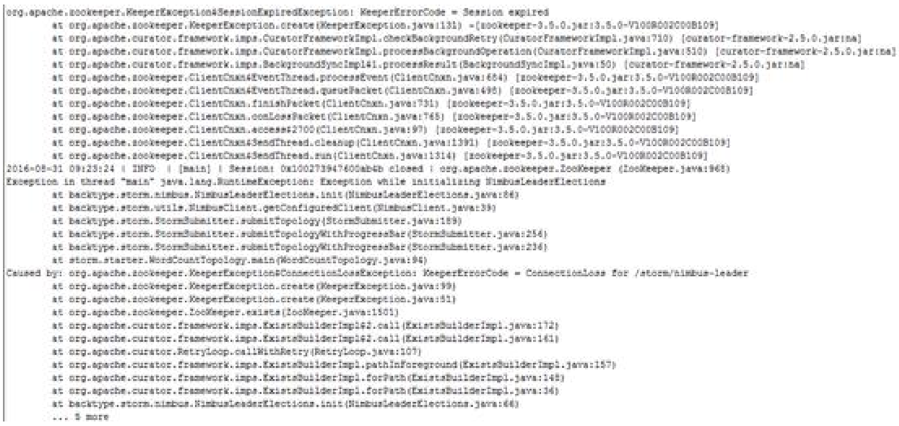
   MRS Manager:

   Log in to MRS Manager. On the MRS Manager page, choose **Services** > **Storm** to check the status of Storm. The status is **Good**, and the monitoring metrics are correctly displayed.

   FusionInsight Manager:

   For MRS 3.*x* or later: Log in to FusionInsight Manager. Choose **Cluster** > **Services** > **Storm** to check the status of Storm. It is found that the status is **Good** and the monitoring metrics are correctly displayed.

2. Check the submission logs of the client. The logs contain "KeeperExceptionSessionExpireException".

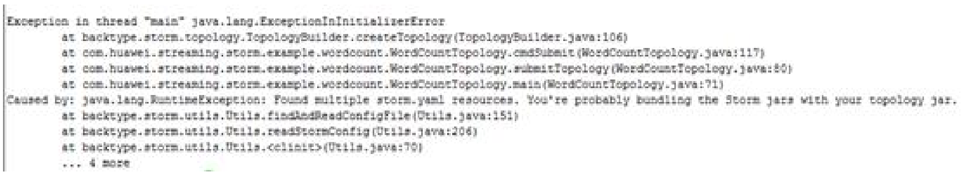


```
org.apache.zookeeper.KeeperException$SessionExpiredException: KeeperErrorCode = Session expired
        at org.apache.zookeeper.KeeperException.create(KeeperException.java:131) ~[zookeeper-3.5.0.jar:3.5.0-V100R002C00B109]
        at org.apache.curator.framework.imps.CuratorFrameworkImpl.checkBackgroundRetry(CuratorFrameworkImpl.java:710) [curator-framework-2.5.0.jar:na]
        at org.apache.curator.framework.imps.CuratorFrameworkImpl.processBackgroundOperation(CuratorFrameworkImpl.java:510) [curator-framework-2.5.0.jar:na]
        at org.apache.curator.framework.imps.BackgroundSyncImpl$1.processResult(BackgroundSyncImpl.java:50) [curator-framework-2.5.0.jar:na]
        at org.apache.zookeeper.ClientCnxn$EventThread.processEvent(ClientCnxn.java:684) [zookeeper-3.5.0.jar:3.5.0-V100R002C00B109]
        at org.apache.zookeeper.ClientCnxn$EventThread.queuePacket(ClientCnxn.java:496) [zookeeper-3.5.0.jar:3.5.0-V100R002C00B109]
        at org.apache.zookeeper.ClientCnxn.finishPacket(ClientCnxn.java:731) [zookeeper-3.5.0.jar:3.5.0-V100R002C00B109]
        at org.apache.zookeeper.ClientCnxn.conLossPacket(ClientCnxn.java:765) [zookeeper-3.5.0.jar:3.5.0-V100R002C00B109]
        at org.apache.zookeeper.ClientCnxn.access$2700(ClientCnxn.java:97) [zookeeper-3.5.0.jar:3.5.0-V100R002C00B109]
        at org.apache.zookeeper.ClientCnxn$SendThread.cleanup(ClientCnxn.java:1391) [zookeeper-3.5.0.jar:3.5.0-V100R002C00B109]
        at org.apache.zookeeper.ClientCnxn$SendThread.run(ClientCnxn.java:1314) [zookeeper-3.5.0.jar:3.5.0-V100R002C00B109]
2016-08-31 09:23:24 | INFO  | [main] | Session: 0x100273947400ab4b closed | org.apache.zookeeper.ZooKeeper (ZooKeeper.java:968)
Exception in thread "main" java.lang.RuntimeException: Exception while initializing NimbusLeaderElections
        at backtype.storm.nimbus.NimbusLeaderElections.init(NimbusLeaderElections.java:66)
        at backtype.storm.utils.NimbusClient.getConfiguredClient(NimbusClient.java:39)
        at backtype.storm.StormSubmitter.submitTopology(StormSubmitter.java:189)
        at backtype.storm.StormSubmitter.submitTopologyWithProgressBar(StormSubmitter.java:254)
        at backtype.storm.StormSubmitter.submitTopologyWithProgressBar(StormSubmitter.java:236)
        at storm.starter.WordCountTopology.main(WordCountTopology.java:94)
Caused by: org.apache.zookeeper.KeeperException$ConnectionLossException: KeeperErrorCode = ConnectionLoss for /storm/nimbus-leader
        at org.apache.zookeeper.KeeperException.create(KeeperException.java:99)
        at org.apache.zookeeper.KeeperException.create(KeeperException.java:51)
        at org.apache.zookeeper.ZooKeeper.exists(ZooKeeper.java:1501)
        at org.apache.curator.framework.imps.ExistsBuilderImpl$2.call(ExistsBuilderImpl.java:172)
        at org.apache.curator.framework.imps.ExistsBuilderImpl$2.call(ExistsBuilderImpl.java:161)
        at org.apache.curator.RetryLoop.callWithRetry(RetryLoop.java:107)
        at org.apache.curator.framework.imps.ExistsBuilderImpl.pathInForeground(ExistsBuilderImpl.java:157)
        at org.apache.curator.framework.imps.ExistsBuilderImpl.forPath(ExistsBuilderImpl.java:148)
        at org.apache.curator.framework.imps.ExistsBuilderImpl.forPath(ExistsBuilderImpl.java:36)
        at backtype.storm.nimbus.NimbusLeaderElections.init(NimbusLeaderElections.java:66)
        ... 5 more
```

The preceding error occurs because security authentication is not performed before the topology is submitted or the TGT expires after authentication.

For details about the solution, see **Step 1**.

3. Check the client submission log. It is found that the "ExceptionInIntializerError" exception information is printed, and the message "Found multiple storm.yaml resources" is displayed. The following is an example:



```
Exception in thread "main" java.lang.ExceptionInInitializerError
        at backtype.storm.topology.TopologyBuilder.createTopology(TopologyBuilder.java:106)
        at com.huawei.streaming.storm.example.wordcount.WordCountTopology.cmdSubmit(WordCountTopology.java:117)
        at com.huawei.streaming.storm.example.wordcount.WordCountTopology.submitTopology(WordCountTopology.java:80)
        at com.huawei.streaming.storm.example.wordcount.WordCountTopology.main(WordCountTopology.java:71)
Caused by: java.lang.RuntimeException: Found multiple storm.yaml resources. You're probably bundling the Storm jars with your topology jar.
        at backtype.storm.utils.Utils.findAndReadConfigFile(Utils.java:151)
        at backtype.storm.utils.Utils.readStormConfig(Utils.java:206)
        at backtype.storm.utils.Utils.<clinit>(Utils.java:70)
        ... 4 more
```

This error occurs because the **storm.yaml** file in the service JAR package conflicts with that on the server.

For details about the solution, see **Step 2**.

4. If the fault is not caused by the preceding reasons, see **Topology Submission Fails and the Message "Failed to check principle for keytab" Is Displayed**.

Solution

Step 1 An authentication error occurs.

1. Log in to the node where the client resides and switch to the client directory.
2. Run the following command to submit the task again: (Replace the service JAR package and topology based on the site requirements.)

   source bigdata_env

   kinit *Username*

   storm jar storm-starter-topologies-0.10.0.jar storm.starter.WordCountTopology test

Step 2 The topology package is abnormal.

Check the service JAR package, delete the **storm.yaml** file from the service JAR package, and submit the task again.

----End

15.18.3 Topology Submission Fails and the Message "Failed to check principle for keytab" Is Displayed

Symptom

An MRS streaming cluster in security mode is installed, and ZooKeeper, Storm, and Kafka are installed in the cluster.

When a topology is defined to access components such as HDFS and HBase and the topology fails to be submitted using client commands.

Possible Causes

- The submitted topology does not contain the keytab file of the user.

- The keytab file contained in the submitted topology is inconsistent with the user who submits the topology.
- The **user.keytab** file exists in the **/tmp** directory on the client, and the owner is not the running user.

Cause Analysis

1. Check the logs. Error information "Can not found user.keytab in storm.jar" is found. Details are as follows:

```
[main] INFO  b.s.StormSubmitter - Get principle for stream@HADOOP.COM success
[main] ERROR b.s.StormSubmitter - Can not found user.keytab in storm.jar.
Exception in thread "main" java.lang.RuntimeException: Failed to check principle for keytab
at backtype.storm.StormSubmitter.submitTopologyAs(StormSubmitter.java:219)
at backtype.storm.StormSubmitter.submitTopology(StormSubmitter.java:292)
at backtype.storm.StormSubmitter.submitTopology(StormSubmitter.java:176)
at com.xxx.streaming.storm.example.hbase.SimpleHBaseTopology.main(SimpleHBaseTopology.java:77)
```

   Check the JAR file of the submitted topology. It is found that the keytab file is not contained.

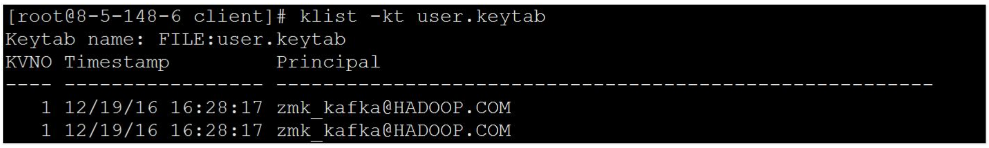
2. Check the logs. Error information "The submit user is invalid,the principle is" is found. Details are as follows:

```
[main] INFO  b.s.StormSubmitter - Get principle for stream@HADOOP.COM success
[main] WARN  b.s.s.a.k.ClientCallbackHandler - Could not login: the client is being asked for a password, but the  client code does not currently support obtaining a password from the user. Make sure that the client is configured to use a ticket cache (using the JAAS configuration setting 'useTicketCache=true)' and restart the client. If you still get this message after that, the TGT in the ticket cache has expired and must be manually refreshed. To do so, first determine if you are using a password or a keytab. If the former, run kinit in a Unix shell in the environment of the user who is running this client using the command 'kinit <princ>' (where <princ> is the name of the client's Kerberos principal). If the latter, do 'kinit -k -t <keytab> <princ>' (where <princ> is the name of the Kerberos principal, and <keytab> is the location of the keytab file). After manually refreshing your cache, restart this client. If you continue to see this message after manually refreshing your cache, ensure that your KDC host's clock is in sync with this host's clock.
[main] ERROR b.s.StormSubmitter - The submit user is invalid,the principle is : stream@HADOOP.COM
Exception in thread "main" java.lang.RuntimeException: Failed to check principle for keytab
at backtype.storm.StormSubmitter.submitTopologyAs(StormSubmitter.java:219)
at backtype.storm.StormSubmitter.submitTopology(StormSubmitter.java:292)
at backtype.storm.StormSubmitter.submitTopology(StormSubmitter.java:176)
at com.xxx.streaming.storm.example.hbase.SimpleHBaseTopology.main(SimpleHBaseTopology.java:77)
```

   The authenticated user used to submit the topology is **stream**. However, the system displays a message indicating that the submit user is invalid during topology submission, indicating that the internal verification fails.

3. Check the JAR file of the submitted topology. It is found that the keytab file is contained.

   The principal parameter is set to **zmk_kafka** in the **user.keytab** file.



```
[root@8-5-148-6 client]# klist -kt user.keytab
Keytab name: FILE:user.keytab
KVNO Timestamp         Principal
---- ----------------- --------------------------------------------------------
   1 12/19/16 16:28:17 zmk_kafka@HADOOP.COM
   1 12/19/16 16:28:17 zmk_kafka@HADOOP.COM
```

   It is found that the authenticated user does not match the principal in the **user.keytab** file.

4. Check the logs and find the error information "Delete the tmp keytab file failed, the keytab file is:/tmp/user.keytab". The detailed information is as follows:

```
[main] WARN b.s.StormSubmitter - Delete the tmp keytab file failed, the keytab file is : /tmp/user.keytab
[main] ERROR b.s.StormSubmitter - The submit user is invalid,the principle is : hbase1@HADOOP.COM
Exception in thread "main" java.lang.RuntimeException: Failed to check principle for keytab
```

```
at backtype.storm.StormSubmitter.submitTopologyAs(StormSubmitter.java:213)
at backtype.storm.StormSubmitter.submitTopology(StormSubmitter.java:286)
at backtype.storm.StormSubmitter.submitTopology(StormSubmitter.java:170)
at com.touchstone.storm.cmcc.CmccDataHbaseTopology.main(CmccDataHbaseTopology.java:183)
```

Check the **/tmp** directory. It is found that the **user.keytab** file exists and the file owner is not the running user.

Solution

- Ensure that the **user.keytab** file is carried when the topology is submitted.
- Ensure that the user for submitting the topology is the same as that of the **user.keytab** file.
- Delete the **user.keytab** file from the **/tmp** directory.

15.18.4 The Worker Log Is Empty After a Topology Is Submitted

Symptom

After a topology is remotely submitted in Eclipse, the detailed information about the topology cannot be viewed on the Storm web UI, and the Worker node where Bolt and Spout of each topology are located keeps changing. The Worker log is empty.

Possible Causes

The Worker process fails to be started, triggering Nimbus to re-allocate tasks and start the Worker process on other Supervisors. The Worker process continues to restart. As a result, the Worker node keeps changing, and the Worker log is empty. The possible causes of the Worker process startup failure are as follows:

- The submitted JAR package contains the **storm.yaml** file.

  Storm specifies that each classpath can contain only one **storm.yaml** file. If there is more than one **storm.yaml** file, an exception occurs. Use the Storm client to submit the topology. The classpath configuration of the client is different from the classpath configuration of Eclipse. The client automatically loads the JAR package of the user to classpath. As a result, two **storm.yaml** files exist in classpath.
- The initialization of the Worker process takes a long time, which exceeds the Worker startup timeout period set in the Storm cluster. As a result, the Worker process is killed and reallocated.

Troubleshooting Process

1. Use the Storm client to submit the topology and check whether the **storm.yaml** file is duplicate.
2. Repack the JAR file and submit the topology again.
3. Modify the Worker startup timeout parameter in the Storm cluster.

Procedure

Step 1 If the Worker log is empty after the topology is remotely submitted using Eclipse, use the Storm client to submit the JAR package corresponding to the topology and view the prompt message.

For example, if the JAR package contains two **storm.yaml** files in different paths, the following information is displayed:

```
Exception in thread "main" java.lang.ExceptionInInitializerError
  at com.xxx.streaming.storm.example.WordCountTopology.createConf(WordCountTopology.java:132)
  at com.xxx.streaming.storm.example.WordCountTopology.remoteSubmit(WordCountTopology.java:120)
  at com.xxx.streaming.storm.example.WordCountTopology.main(WordCountTopology.java:101)
 Caused by: java.lang.RuntimeException: Found multiple storm.yaml resources. You're probably bundling the
Storm jars with your topology jar. [jar:file:/opt/xxx/fi_client/Streaming/streaming-0.9.2/bin/stormDemo.jar!/
storm.yaml, file:/opt/xxx/fi_client/Streaming/streaming-0.9.2/conf/storm.yaml]
  at backtype.storm.utils.Utils.findAndReadConfigFile(Utils.java:151)
  at backtype.storm.utils.Utils.readStormConfig(Utils.java:206)
  at backtype.storm.utils.Utils.<(Utils.java:70)>
```

Step 2 Compress the JAR package again. Ensure that the package does not contain the **storm.yaml** file and JAR packages related to **log4j** and **slf4j-log4j**.

Step 3 Use IntelliJ IDEA to remotely submit the new JAR package.

Step 4 Check whether the topology details and Worker logs can be viewed on the web UI.

Step 5 On MRS Manager, modify the Worker startup timeout parameter of the Storm cluster (for details about the parameter description, see **Related Information**). Save the modification, and restart the Storm service.

- MRS Manager: Log in to MRS Manager and choose **Services** > **Storm** > **Configuration**.
- FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Storm** > **Configuration**.

Step 6 Submit the JAR package to be run again.

----End

Related Information

1. The **nimbus.task.launch.secs** and **supervisor.worker.start.timeout.secs** parameters indicate the topology startup timeout tolerance of the Nimbus and supervisor, respectively. Generally, the value of **nimbus.task.launch.secs** must be greater than or equal to that of **supervisor.worker.start.timeout.secs**. It is recommended that the value of **nimbus.task.launch.secs** be slightly greater or equal to that of **supervisor.worker.start.timeout.secs**. Otherwise, the task reallocation efficiency will be affected.
   – **nimbus.task.launch.secs**: If the Nimbus does not receive the heartbeat message sent by the topology task within the period specified by this parameter, the Nimbus re-allocates the topology to another supervisor and updates the task information in ZooKeeper. The supervisor reads the task information in ZooKeeper and compares it with the topology started. If the topology does not belong to the supervisor, the supervisor deletes the metadata of the topology, that is, the **/srv/Bigdata/streaming_data/stormdir/supervisor/stormdist/{worker-id}** directory.

- **supervisor.worker.start.timeout.secs**: After the supervisor starts a worker, if no heartbeat message is received from the worker within the period specified by this parameter, the supervisor stops the worker and waits for worker rescheduling. Generally, the value of this parameter is increased when the service startup takes a long time to ensure that the worker can be started successfully.

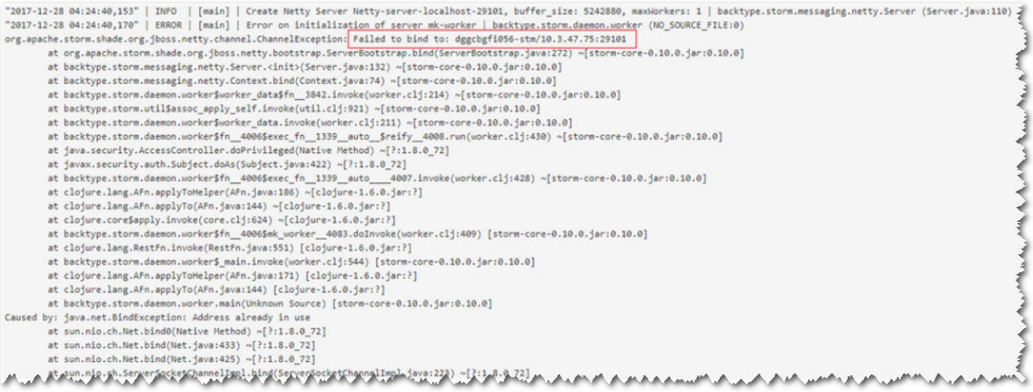
  If the value of **supervisor.worker.start.timeout.secs** is greater than that of **nimbus.task.launch.secs**, the worker is still started before the tolerance time of supervisor ends. However, the Nimbus considers that the service startup times out and allocates the service to another host. The background thread of the supervisor finds that the tasks are inconsistent and deletes the metadata of the topology. As a result, when the worker attempts to read **stormconf.ser** during startup, the file does not exist, and "FileNotFoundException" is thrown.

2. The **nimbus.task.timeout.secs** and **supervisor.worker.timeout.secs** parameters indicate the timeout tolerance time for the Nimbus and supervisor to report heartbeat messages during topology running. Generally, the value of **nimbus.task.timeout.secs** must be slightly greater than or equal to that of **supervisor.worker.timeout.secs**.

15.18.5 Worker Runs Abnormally After a Topology Is Submitted and Error "Failed to bind to:host:ip" Is Displayed

Symptom

After the service topology is submitted, the Worker cannot be started normally. Check the Worker log. The log records "Failed to bind to: host:ip."



```
"2017-12-28 04:24:40,153" | INFO  | [main] | Create Netty Server Netty-server-localhost-29101, buffer_size: 5242880, maxWorkers: 1 | backtype.storm.messaging.netty.Server (Server.java:110)
"2017-12-28 04:24:40,170" | ERROR | [main] | Error on initialization of server mk-worker | backtype.storm.daemon.worker (NO_SOURCE_FILE:0)
org.apache.storm.shade.org.jboss.netty.channel.ChannelException: Failed to bind to: dggcbgf1056-stm/10.3.47.75:29101
        at org.apache.storm.shade.org.jboss.netty.bootstrap.ServerBootstrap.bind(ServerBootstrap.java:272) ~[storm-core-0.10.0.jar:0.10.0]
        at backtype.storm.messaging.netty.Server.<init>(Server.java:132) ~[storm-core-0.10.0.jar:0.10.0]
        at backtype.storm.messaging.netty.Context.bind(Context.java:74) ~[storm-core-0.10.0.jar:0.10.0]
        at backtype.storm.daemon.worker$worker_data$fn__3842.invoke(worker.clj:214) ~[storm-core-0.10.0.jar:0.10.0]
        at backtype.storm.util$assoc_apply_self.invoke(util.clj:921) ~[storm-core-0.10.0.jar:0.10.0]
        at backtype.storm.daemon.worker$worker_data.invoke(worker.clj:211) ~[storm-core-0.10.0.jar:0.10.0]
        at backtype.storm.daemon.worker$fn__4006$exec_fn__1339__auto__$reify__4008.run(worker.clj:430) ~[storm-core-0.10.0.jar:0.10.0]
        at java.security.AccessController.doPrivileged(Native Method) ~[?:1.8.0_72]
        at javax.security.auth.Subject.doAs(Subject.java:422) ~[?:1.8.0_72]
        at backtype.storm.daemon.worker$fn__4006$exec_fn__1339__auto____4007.invoke(worker.clj:428) ~[storm-core-0.10.0.jar:0.10.0]
        at clojure.lang.AFn.applyToHelper(AFn.java:186) ~[clojure-1.6.0.jar:?]
        at clojure.lang.AFn.applyTo(AFn.java:144) ~[clojure-1.6.0.jar:?]
        at clojure.core$apply.invoke(core.clj:624) ~[clojure-1.6.0.jar:?]
        at backtype.storm.daemon.worker$fn__4006$mk_worker__4083.doInvoke(worker.clj:409) [storm-core-0.10.0.jar:0.10.0]
        at clojure.lang.RestFn.invoke(RestFn.java:551) [clojure-1.6.0.jar:?]
        at backtype.storm.daemon.worker$_main.invoke(worker.clj:544) [storm-core-0.10.0.jar:0.10.0]
        at clojure.lang.AFn.applyToHelper(AFn.java:171) [clojure-1.6.0.jar:?]
        at clojure.lang.AFn.applyTo(AFn.java:144) [clojure-1.6.0.jar:?]
        at backtype.storm.daemon.worker.main(Unknown Source) [storm-core-0.10.0.jar:0.10.0]
Caused by: java.net.BindException: Address already in use
        at sun.nio.ch.Net.bind0(Native Method) ~[?:1.8.0_72]
        at sun.nio.ch.Net.bind(Net.java:433) ~[?:1.8.0_72]
        at sun.nio.ch.Net.bind(Net.java:425) ~[?:1.8.0_72]
        at sun.nio.ch.ServerSocketChannelImpl.bind(ServerSocketChannelImpl.java:223) ~[?:1.8.0_72]
```

Possible Causes

The random port range is incorrectly configured.

Troubleshooting Process

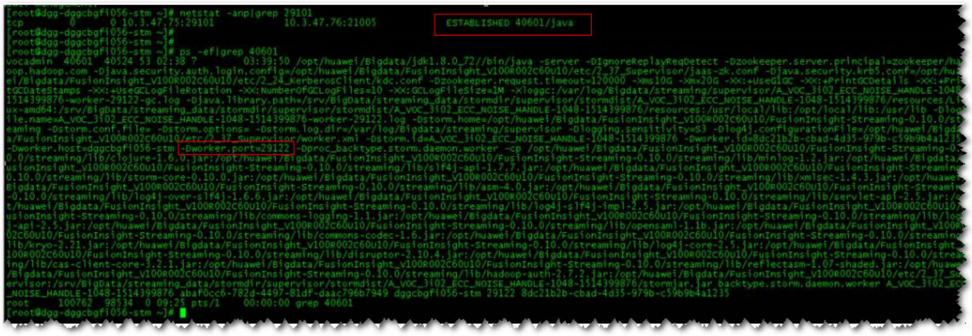
1. Check related information in the Worker log.

2. Check the process information about the bond port.

3. Check the random port range.

Cause Analysis

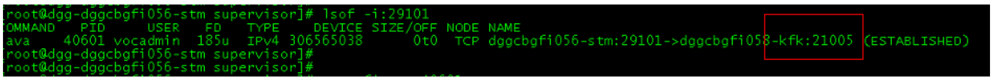
1. Use SSH to log in to the host where the Worker fails to be started and run the **netstat -anp | grep** <*port*> command to check the ID of the process that occupies the port. In the preceding command, change *port* to the actual port number.
2. Run the **ps -ef | grep** <*pid*> command to view process details. In the command, *pid* indicates the actual process ID.



```
[root@dgg-dggcbgfi056-stm ~]# netstat -anp|grep 29101
tcp        0      0 10.3.47.75:29101            10.3.47.76:21005            ESTABLISHED 40601/java
[root@dgg-dggcbgfi056-stm ~]#
[root@dgg-dggcbgfi056-stm ~]#
[root@dgg-dggcbgfi056-stm ~]# ps -ef|grep 40601
vocadmin  40601  40524 53 02:38 ?        03:39:50 /opt/huawei/Bigdata/jdk1.8.0_72//bin/java -server -DIgnoreReplayReqDetect -Dzookeeper.server.principal=zookeeper/ha
oop.hadoop.com -Djava.security.auth.login.config=/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/etc/2_37_Supervisor/jaas-zk.conf -Djava.security.krb5.conf=/opt/hu
ei/Bigdata/FusionInsight_V100R002C60U10/etc/2_24_KerberosClient/kdc.conf -Dzookeeper.request.timeout=120000 -Xms10G -Xmx20G -XX:+UseG1GC -XX:+PrintGCDetails -XX:+Pri
tGCDateStamps -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=1M -Xloggc:/var/log/Bigdata/streaming/supervisor/A_VOC_3i02_ECC_NOISE_HANDLE-1048
1514399876-worker-29122-gc.log -Djava.library.path=/srv/BigData/streaming_data/stormdir/supervisor/stormdist/A_VOC_3i02_ECC_NOISE_HANDLE-1048-1514399876/resources/Li
ux-amd64:/srv/BigData/streaming_data/stormdir/supervisor/stormdist/A_VOC_3i02_ECC_NOISE_HANDLE-1048-1514399876/resources:/usr/local/lib:/opt/local/lib:/usr/lib -Dlog
ile.name=A_VOC_3i02_ECC_NOISE_HANDLE-1048-1514399876-worker-29122.log -Dstorm.home=/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/st
eaming -Dstorm.conf.file= -Dstorm.options= -Dstorm.log.dir=/var/log/Bigdata/streaming/supervisor -Dlogging.sensitivity=S3 -Dlog4j.configurationFile=/opt/huawei/Bigda
a/FusionInsight_V100R002C60U10/etc/2_37_Supervisor/worker.xml -Dstorm.id=A_VOC_3i02_ECC_NOISE_HANDLE-1048-1514399876 -Dworker.id=8dc21b2b-cbad-4d35-979b-c59b9b4a1235
-Dworker.host=dggcbgfi056-stm -Dworker.port=29122 -Dproc_backtype.storm.daemon.worker -cp /opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.
0.0/streaming/lib/clojure-1.6.0.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/streaming/lib/minlog-1.2.jar:/opt/huawei/Bigdata/
usionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/streaming/lib/slf4j-api-1.7.7.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming
0.10.0/streaming/lib/storm-core-0.10.0.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/streaming/lib/xmlsec-1.4.3.jar:/opt/huawei
Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/streaming/lib/asm-4.0.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streamin
-0.10.0/streaming/lib/log4j-over-slf4j-1.6.6.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/streaming/lib/servlet-api-2.5.jar:/o
t/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/streaming/lib/log4j-slf4j-impl-2.5.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/
usionInsight-Streaming-0.10.0/streaming/lib/commons-logging-1.1.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/streaming/lib/log
j-api-2.5.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/streaming/lib/opensaml-1.1b.jar:/opt/huawei/Bigdata/FusionInsight_V100R
02C60U10/FusionInsight-Streaming-0.10.0/streaming/lib/commons-codec-1.6.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/streaming
lib/kryo-2.21.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/streaming/lib/log4j-core-2.5.jar:/opt/huawei/Bigdata/FusionInsight_
100R002C60U10/FusionInsight-Streaming-0.10.0/streaming/lib/disruptor-2.10.4.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/strea
ing/lib/cas-client-core-3.2.1.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/streaming/lib/reflectasm-1.07-shaded.jar:/opt/huaw
/Bigdata/FusionInsight_V100R002C60U10/FusionInsight-Streaming-0.10.0/streaming/lib/hadoop-auth-2.7.2.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U10/etc/2_37_Su
ervisor:/srv/BigData/streaming_data/stormdir/supervisor/stormdist/A_VOC_3i02_ECC_NOISE_HANDLE-1048-1514399876/stormjar.jar backtype.storm.daemon.worker A_VOC_3i02_EC
_NOISE_HANDLE-1048-1514399876 abaf0cc6-782d-4497-81df-daac796b7949 dggcbgfi056-stm 29122 8dc21b2b-cbad-4d35-979b-c59b9b4a1235
root     100762  98534  0 09:25 pts/1    00:00:00 grep 40601
[root@dgg-dggcbgfi056-stm ~]#
```

It is found that the worker process occupies the port. This process is another topology service process. According to the process details, port 29122 is allocated to the process.

3. Run the **lsof -i:**<*port*> command to view connection details. In the preceding command, change *port* to the actual port number.

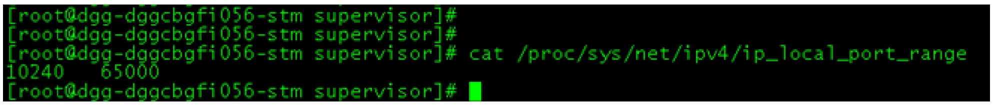


```
root@dgg-dggcbgfi056-stm supervisor]# lsof -i:29101
OMMAND    PID     USER   FD   TYPE    DEVICE SIZE/OFF NODE NAME
ava     40601 vocadmin  185u  IPv4 306565038      0t0  TCP dggcbgfi056-stm:29101->dggcbgfi058-kfk:21005 (ESTABLISHED)
root@dgg-dggcbgfi056-stm supervisor]#
root@dgg-dggcbgfi056-stm supervisor]#
```

It is found that port 29101 connects to port 21005 of the peer end, and port 21005 is the Kafka server port.

It indicates that the service layer connects to Kafka to obtain messages as a client. Service ports are allocated based on the random port range of the OS.

4. Run the **cat /proc/sys/net/ipv4/ip_local_port_range** command to check the random port range.



```
[root@dgg-dggcbgfi056-stm supervisor]#
[root@dgg-dggcbgfi056-stm supervisor]#
[root@dgg-dggcbgfi056-stm supervisor]# cat /proc/sys/net/ipv4/ip_local_port_range
10240   65000
[root@dgg-dggcbgfi056-stm supervisor]#
```

5. It is found that the random port range is too large and conflicts with the service port range of MRS.

NOTE

The MRS service port number ranges from 20000 to 30000.

Procedure

Step 1 Modify the random port range.

```
vi /proc/sys/net/ipv4/ip_local_port_range
32768 61000
```

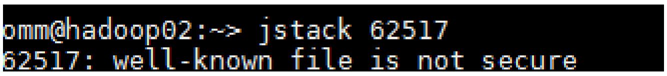
Step 2 Stop the service process that occupies the service port to release the port. (Stop the service topology.)

----End

15.18.6 "well-known file is not secure" Is Displayed When the jstack Command Is Used to Check the Process Stack

Symptom

Run the **jstack** command to check the process stack information. The error message "well-known file is not secure" is displayed.



```
omm@hadoop02:~> jstack 62517
62517: well-known file is not secure
```

Cause Analysis

1. The user running the **jstack** command is inconsistent with the user submitting the process for viewing the pid information.
2. Storm uses the feature of differentiating users for implementing tasks. When the worker process is started, the process UID and GID are changed to the user submitting the task and ficommon. This way, logviewer can access logs of the worker process and only log file permission 640 is open. After the user is changed, the **jstack** and **jmap** commands fail to be executed for the worker process, because the default GID of the user is not ficommon. You need to run the ldap command to change the user GID to 9998 (ficommon).

Solution

You can use either of the following two methods to resolve the problem:

Method 1: View the process stack on the native Storm page.

Step 1 Log in to the native Storm page.

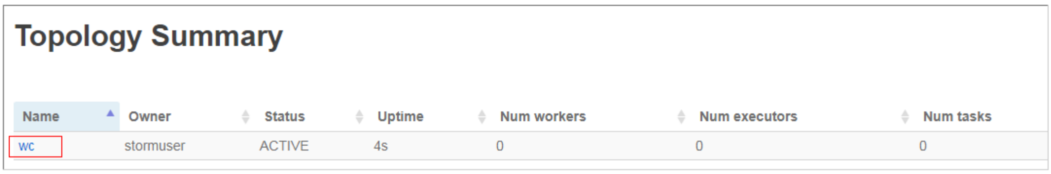
MRS Manager:

1. Access MRS Manager.
2. Choose **Services** > **Storm**. In **Storm WebUI** of **Storm Summary**, click any UI link to access the Storm WebUI.

FusionInsight Manager:

1. Log in to FusionInsight Manager.
2. On Manager, choose **Cluster > Service** > **Storm**. On the **Storm WebUI** page of **Overview**, click any UI link to open the Storm WebUI.

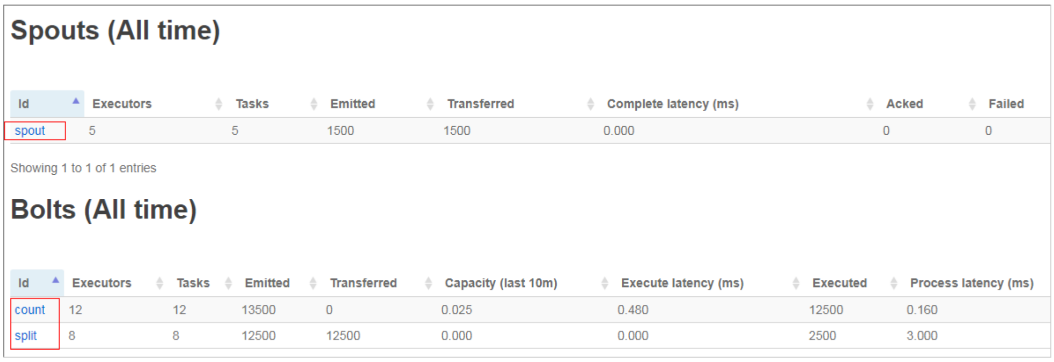
Step 2 Select the topology to be viewed.



Topology Summary

Name	Owner	Status	Uptime	Num workers	Num executors	Num tasks
wc	stormuser	ACTIVE	4s	0	0	0

Step 3 Select the spout or bolt to be viewed.



Spouts (All time)

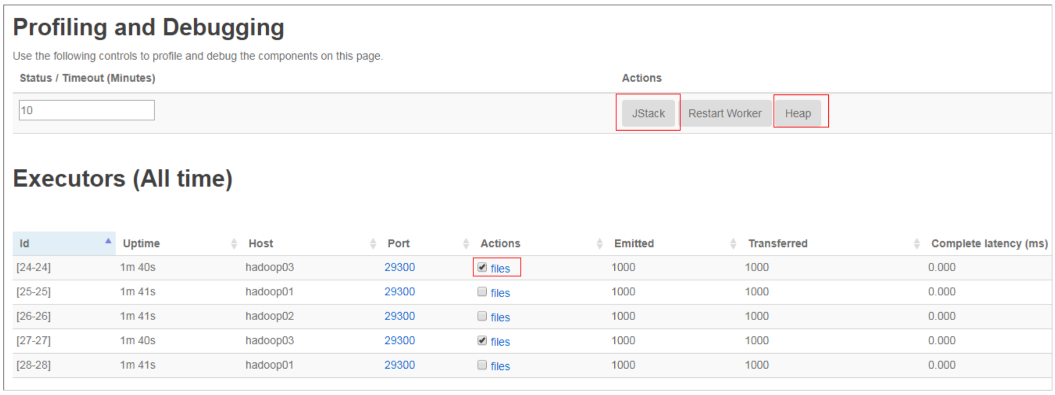
Id	Executors	Tasks	Emitted	Transferred	Complete latency (ms)	Acked	Failed
spout	5	5	1500	1500	0.000	0	0

Showing 1 to 1 of 1 entries

Bolts (All time)

Id	Executors	Tasks	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)
count	12	12	13500	0	0.025	0.480	12500	0.160
split	8	8	12500	12500	0.000	0.000	2500	3.000

Step 4 Select the log file of the node to be viewed, and then click **JStack** or **Heap**. **JStack** corresponds to the stack information, and **Heap** corresponds to the heap information.



Profiling and Debugging

Use the following controls to profile and debug the components on this page.

Status / Timeout (Minutes)	Actions
10	JStack Restart Worker Heap

Executors (All time)

Id	Uptime	Host	Port	Actions	Emitted	Transferred	Complete latency (ms)
[24-24]	1m 40s	hadoop03	29300	☑ files	1000	1000	0.000
[25-25]	1m 41s	hadoop01	29300	□ files	1000	1000	0.000
[26-26]	1m 41s	hadoop02	29300	□ files	1000	1000	0.000
[27-27]	1m 40s	hadoop03	29300	☑ files	1000	1000	0.000
[28-28]	1m 41s	hadoop01	29300	□ files	1000	1000	0.000

----End

Method 2: View the process stack by modifying user-defined parameters.

Step 1 Access the Storm parameter configuration page.

MRS Manager: Log in to MRS Manager, choose **Services** > **Storm** > **Service Configuration**, and select **All** from the **Type** drop-down list.

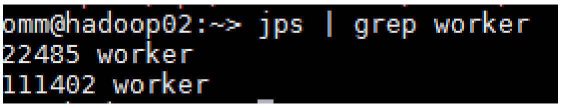
Operation on FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > **Services** > **Yarn** > **Configurations** > **All Configurations**.

Step 2 In the navigation tree on the left, choose **supervisor** > **Customize** and add the variable **supervisor.run.worker.as.user=false**.

Step 3 Click **Save Configuration** and select **Restart the affected services or instances**. Click **OK** to restart the services.

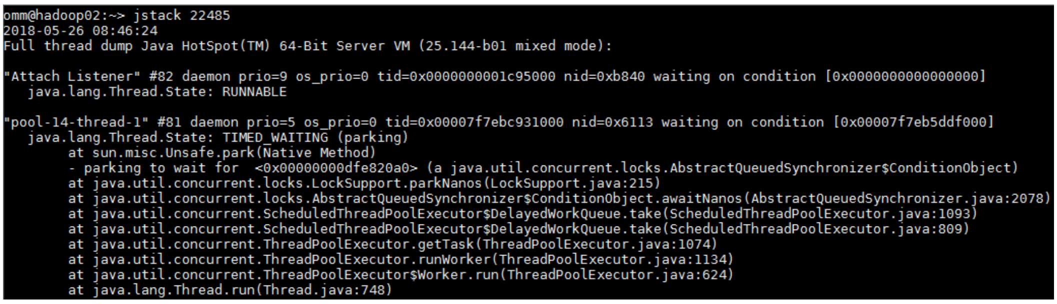
Step 4 Submit the topology again.

Step 5 Switch to the **omm** user on the background node and run the **jps** command to view the PID of the worker process.



```
omm@hadoop02:~> jps | grep worker
22485 worker
111402 worker
```

Step 6 Run the **jstack pid** command to view the jstack information.



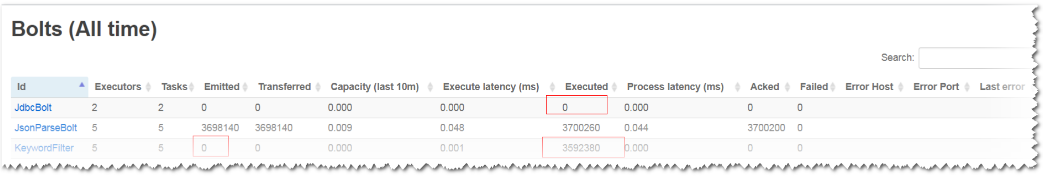
```
omm@hadoop02:~> jstack 22485
2018-05-26 08:46:24
Full thread dump Java HotSpot(TM) 64-Bit Server VM (25.144-b01 mixed mode):

"Attach Listener" #82 daemon prio=9 os_prio=0 tid=0x0000000001c95000 nid=0xb840 waiting on condition [0x0000000000000000]
   java.lang.Thread.State: RUNNABLE

"pool-14-thread-1" #81 daemon prio=5 os_prio=0 tid=0x00007f7ebc931000 nid=0x6113 waiting on condition [0x00007f7eb5ddf000]
   java.lang.Thread.State: TIMED_WAITING (parking)
        at sun.misc.Unsafe.park(Native Method)
        - parking to wait for  <0x00000000dfe820a0> (a java.util.concurrent.locks.AbstractQueuedSynchronizer$ConditionObject)
        at java.util.concurrent.locks.LockSupport.parkNanos(LockSupport.java:215)
        at java.util.concurrent.locks.AbstractQueuedSynchronizer$ConditionObject.awaitNanos(AbstractQueuedSynchronizer.java:2078)
        at java.util.concurrent.ScheduledThreadPoolExecutor$DelayedWorkQueue.take(ScheduledThreadPoolExecutor.java:1093)
        at java.util.concurrent.ScheduledThreadPoolExecutor$DelayedWorkQueue.take(ScheduledThreadPoolExecutor.java:809)
        at java.util.concurrent.ThreadPoolExecutor.getTask(ThreadPoolExecutor.java:1074)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1134)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
        at java.lang.Thread.run(Thread.java:748)
```

----End

15.18.7 When the Storm-JDBC plug-in is used to develop Oracle write Bolts, data cannot be written into the Bolts.

Symptom

When the Storm-JDBC plug-in is used to develop Oracle write Bolts, the Oracle database can be connected, but data cannot be written to the Oracle database.



Bolts (All time)

Search:

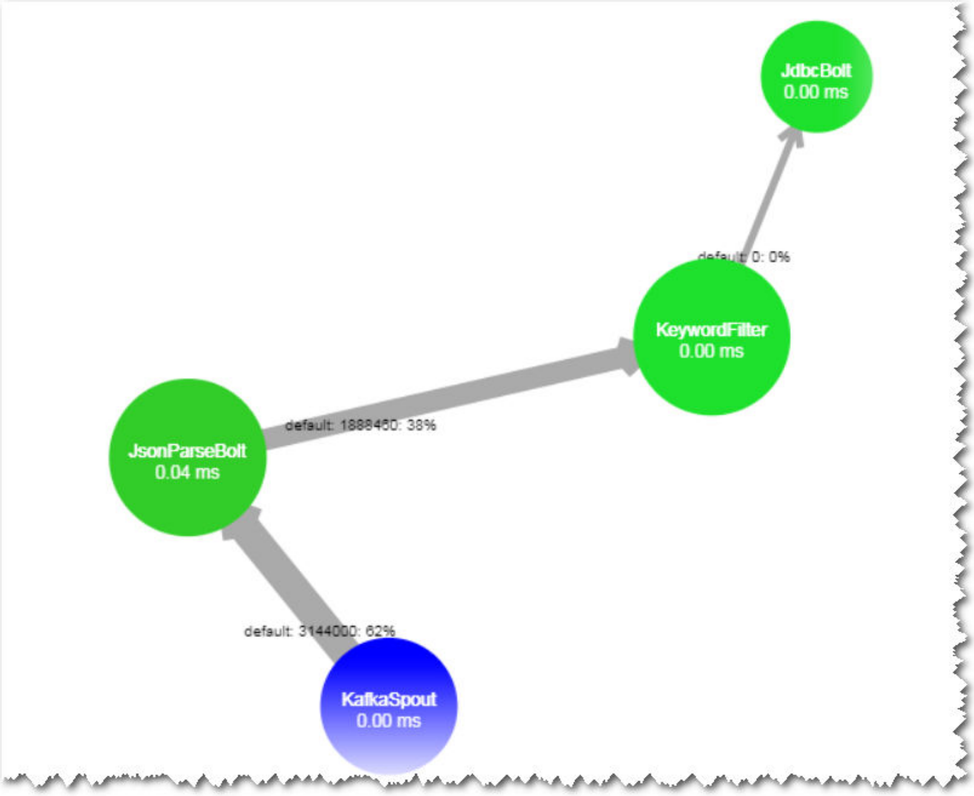
Id	Executors	Tasks	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed	Error Host	Error Port	Last error
JdbcBolt	2	2	0	0	0.000	0.000	0	0.000	0	0			
JsonParseBolt	5	5	3698140	3698140	0.009	0.048	3700260	0.044	3700200	0			
KeywordFilter	5	5	0	0	0.000	0.001	3592380	0.000	0	0			

Possible Causes

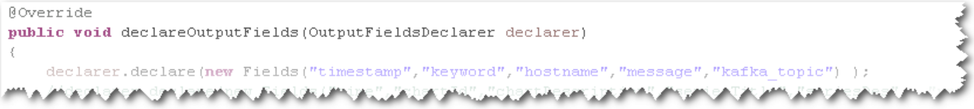
- The topology definition is incorrect.
- The definition of the database table result is incorrect.

Cause Analysis

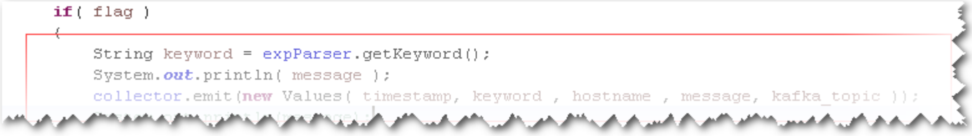
1. On the Storm web UI, check the DAG of the topology. The DAG is consistent with the topology definition.



2. The definition of the KeyWordFilter Bolt is consistent with the **expParser** field.

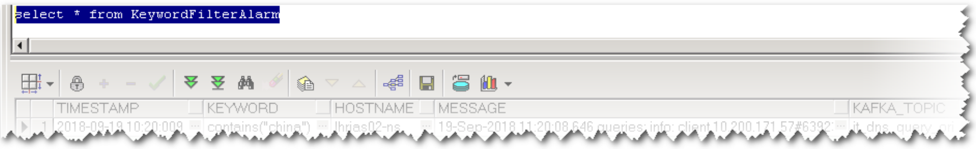


```
@Override
public void declareOutputFields(OutputFieldsDeclarer declarer)
{
    declarer.declare(new Fields("timestamp","keyword","hostname","message","kafka_topic") );
```



```
    if( flag )
    {
        String keyword = expParser.getKeyword();
        System.out.println( message );
        collector.emit(new Values( timestamp, keyword , hostname , message, kafka_topic ));
```

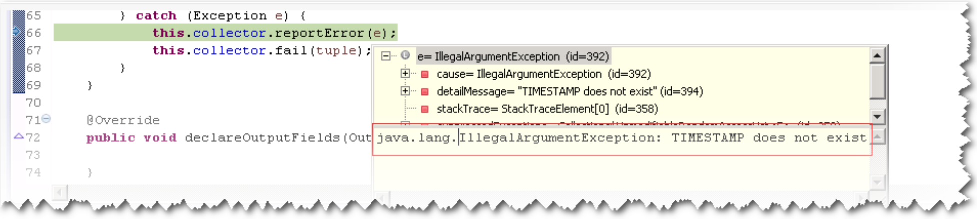
3. View the table definition in the Oracle database. The field name is in uppercase, which is inconsistent with flow definition field name.



```
select * from KeywordFilterAlarm
```

	TIMESTAMP	KEYWORD	HOSTNAME	MESSAGE	KAFKA_TOPIC
1	2018-09-19 10:20:009	contains("china")	lhrias02-ns	19-Sep-2018 11:20:08.646 queries: info: client 10.200.171.57#6392:	it_dns_query_ori

4. When the execute method is debugged independently, it is found that the thrown field does not exist.



```
        } catch (Exception e) {
            this.collector.reportError(e);
            this.collector.fail(tuple);
        }
    }

    @Override
    public void declareOutputFields(Out

    }
```

```
e= IllegalArgumentException  (id=392)
    cause= IllegalArgumentException  (id=392)
    detailMessage= "TIMESTAMP does not exist" (id=394)
    stackTrace= StackTraceElement[0]  (id=358)
java.lang.IllegalArgumentException: TIMESTAMP does not exist
```

Procedure

The field name of the stream definition is changed to uppercase letters, which is the same as that defined in the database table.

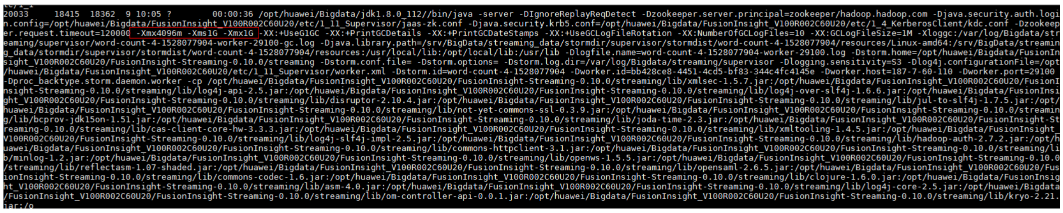
15.18.8 The GC Parameter Configured for the Service Topology Does Not Take Effect

Symptom

The **topology.worker.childopts** parameter in the service topology code does not take effect. The key log is as follows:

```
[main] INFO b.s.StormSubmitter - Uploading topology jar /opt/jar/example.jar to assigned location: /srv/
BigData/streaming/stormdir/nimbus/inbox/stormjar-8d3b778d-69ea-4fbe-ba88-01aa2036d753.jar
Start uploading file '/opt/jar/example.jar' to '/srv/BigData/streaming/stormdir/nimbus/inbox/
stormjar-8d3b778d-69ea-4fbe-ba88-01aa2036d753.jar' (65574612 bytes)
[==================================================] 65574612 / 65574612
File '/opt/jar/example.jar' uploaded to '/srv/BigData/streaming/stormdir/nimbus/inbox/
stormjar-8d3b778d-69ea-4fbe-ba88-01aa2036d753.jar' (65574612 bytes)
[main] INFO b.s.StormSubmitter - Successfully uploaded topology jar to assigned location: /srv/BigData/
streaming/stormdir/nimbus/inbox/stormjar-8d3b778d-69ea-4fbe-ba88-01aa2036d753.jar
[main] INFO b.s.StormSubmitter - Submitting topology word-count in distributed mode with conf
{"topology.worker.childopts":"-
Xmx4096m","storm.zookeeper.topology.auth.scheme":"digest","storm.zookeeper.topology.auth.payload":"-59
15065013522446406:-6421330379815193999","topology.workers":1}
[main] INFO b.s.StormSubmitter - Finished submitting topology: word-count
```

The following worker process information is displayed after the **ps -ef | grep worker** command is executed:



```
tc/1_1
20033     18415  18362  9 10:05 ?        00:00:36 /opt/huawei/Bigdata/jdk1.8.0_112//bin/java -server -DIgnoreReplayReqDetect -Dzookeeper.server.principal=zookeeper/hadoop.hadoop.com -Djava.security.auth.logi
n.config=/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/etc/1_11_Supervisor/jaas-zk.conf -Djava.security.krb5.conf=/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/etc/1_4_KerberosClient/kdc.conf -Dzookeep
er.request.timeout=120000 -Xmx4096m -Xms1G -Xmx1G -XX:+UseG1GC -XX:+PrintGCDetails -XX:+PrintGCDateStamps -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=1M -Xloggc:/var/log/Bigdata/str
eaming/supervisor/word-count-4-1528077904-worker-29100-gc.log -Djava.library.path=/srv/BigData/streaming_data/stormdir/supervisor/stormdist/word-count-4-1528077904/resources/Linux-amd64:/srv/BigData/streamin
g_data/stormdir/supervisor/stormdist/word-count-4-1528077904/resources:/usr/local/lib:/opt/local/lib:/usr/lib -Dlogfile.name=word-count-4-1528077904-worker-29100.log -Dstorm.home=/opt/huawei/Bigdata/FusionIn
sight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming -Dstorm.conf.file= -Dstorm.options= -Dstorm.log.dir=/var/log/Bigdata/streaming/supervisor -Dlogging.sensitivity=S3 -Dlog4j.configurationFile=/opt
/huawei/Bigdata/FusionInsight_V100R002C60U20/etc/1_11_Supervisor/worker.xml -Dstorm.id=word-count-4-1528077904 -Dworker.id=bb428ce8-4451-4cd5-bf83-344c4fc4145e -Dworker.host=187-7-60-110 -Dworker.port=29100
-Dproc_backtype.storm.daemon.worker -cp /opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/xmlsec-1.5.7.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionI
nsight-Streaming-0.10.0/streaming/lib/log4j-api-2.5.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/log4j-over-slf4j-1.6.6.jar:/opt/huawei/Bigdata/FusionInsi
ght_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/disruptor-2.10.4.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/jul-to-slf4j-1.7.5.jar:/opt/
huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/not-yet-commons-ssl-0.3.9.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streamin
g/lib/bcprov-jdk15on-1.51.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/joda-time-2.3.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-St
reaming-0.10.0/streaming/lib/cas-client-core-hw-3.3.3.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/xmltooling-1.4.5.jar:/opt/huawei/Bigdata/FusionInsight_
V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/log4j-slf4j-impl-2.5.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/hadoop-auth-2.7.2.jar:/opt/h
uawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/commons-httpclient-3.1.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/li
b/minlog-1.2.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/openws-1.5.5.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0
/streaming/lib/reflectasm-1.07-shaded.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/opensaml-2.6.5.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/Fus
ionInsight-Streaming-0.10.0/streaming/lib/commons-codec-1.6.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/clojure-1.6.0.jar:/opt/huawei/Bigdata/FusionInsig
ht_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/asm-4.0.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/log4j-core-2.5.jar:/opt/huawei/Bigdata
/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/om-controller-api-0.0.1.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/kryo-2.21.
jar:/o
```

Cause Analysis

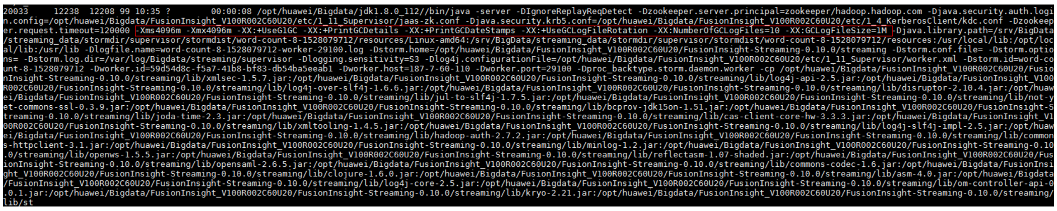
1. **topology.worker.gc.childopts**, **topology.worker.childopts**, and **worker.gc.childopts** (server parameters) have priorities: **topology.worker.gc.childopts** > **worker.gc.childopts** > **topology.worker.childopts**.
2. If the client parameter **topology.worker.childopts** is set, this parameter and the server parameter **worker.gc.childopts** are configured together. However, for two same parameters, one of them will be overwritten by the other parameter after it. Take parameter **-Xmx**, as shown in the red box of the preceding figure, as an example, parameter **-Xmx1G** overwrites **-Xmx4096m**.
3. If parameter **topology.worker.gc.childopts** is configured on the client, the parameter **worker.gc.childopts** on the server will be replaced.

Solution

Step 1 If you want to modify the JVM parameter of the topology, you can directly modify the **topology.worker.gc.childopts** parameter in the command or modify the parameter on the server. When **topology.worker.gc.childopts** is set to **-Xms4096m -Xmx4096m -XX:+UseG1GC -XX:+PrintGCDetails -XX:+PrintGCDateStamps -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=1M**:

```
[main-SendThread(10.7.61.88:2181)] INFO o.a.s.s.o.a.z.ClientCnxn - Socket connection established, initiating
session, client: /10.7.61.88:44694, server: 10.7.61.88/10.7.61.88:2181
[main-SendThread(10.7.61.88:2181)] INFO o.a.s.s.o.a.z.ClientCnxn - Session establishment complete on
server 10.7.61.88/10.7.61.88:2181, sessionid = 0x16037a6e5f092575, negotiated timeout = 40000
[main-EventThread] INFO o.a.s.s.o.a.c.f.s.ConnectionStateManager - State change: CONNECTED
[main] INFO b.s.u.StormBoundedExponentialBackoffRetry - The baseSleepTimeMs [1000] the
maxSleepTimeMs [1000] the maxRetries [1]
[main] INFO o.a.s.s.o.a.z.Login - successfully logged in.
[main-EventThread] INFO o.a.s.s.o.a.z.ClientCnxn - EventThread shut down for session: 0x16037a6e5f092575
[main] INFO o.a.s.s.o.a.z.ZooKeeper - Session: 0x16037a6e5f092575 closed
[main] INFO b.s.StormSubmitter - Uploading topology jar /opt/jar/example.jar to assigned location: /srv/
BigData/streaming/stormdir/nimbus/inbox/stormjar-86855b6b-133e-478d-b415-fa96e63e553f.jar
Start uploading file '/opt/jar/example.jar' to '/srv/BigData/streaming/stormdir/nimbus/inbox/
stormjar-86855b6b-133e-478d-b415-fa96e63e553f.jar' (74143745 bytes)
[==================================================] 74143745 / 74143745
File '/opt/jar/example.jar' uploaded to '/srv/BigData/streaming/stormdir/nimbus/inbox/
stormjar-86855b6b-133e-478d-b415-fa96e63e553f.jar' (74143745 bytes)
[main] INFO b.s.StormSubmitter - Successfully uploaded topology jar to assigned location: /srv/BigData/
streaming/stormdir/nimbus/inbox/stormjar-86855b6b-133e-478d-b415-fa96e63e553f.jar
[main] INFO b.s.StormSubmitter - Submitting topology word-count in distributed mode with conf
{"storm.zookeeper.topology.auth.scheme":"digest","storm.zookeeper.topology.auth.payload":"-736000280424
1426074:-6868950379453400421","topology.worker.gc.childopts":"-Xms4096m -Xmx4096m -XX:+UseG1GC -
XX:+PrintGCDetails -XX:+PrintGCDateStamps -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -
XX:GCLogFileSize=1M","topology.workers":1}
[main] INFO b.s.StormSubmitter - Finished submitting topology: word-count
```

Step 2 Run the **ps -ef | grep worker** command to view the worker process information:



```
20033     12238  12208 99 10:35 ?        00:00:08 /opt/huawei/Bigdata/jdk1.8.0_112//bin/java -server -DIgnoreReplayReqDetect -Dzookeeper.server.principal=zookeeper/hadoop.hadoop.com -Djava.security.auth.logi
n.config=/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/etc/1_11_Supervisor/jaas-zk.conf -Djava.security.krb5.conf=/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/etc/1_4_KerberosClient/kdc.conf -Dzookeep
er.request.timeout=120000 -Xms4096m -Xmx4096m -XX:+UseG1GC -XX:+PrintGCDetails -XX:+PrintGCDateStamps -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=1M -Djava.library.path=/srv/BigData
/streaming_data/stormdir/supervisor/stormdist/word-count-8-1528079712/resources/Linux-amd64:/srv/BigData/streaming_data/stormdir/supervisor/stormdist/word-count-8-1528079712/resources:/usr/local/lib:/opt/loc
al/lib:/usr/lib -Dlogfile.name=word-count-8-1528079712-worker-29100.log -Dstorm.home=/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming -Dstorm.conf.file= -Dstorm.optio
ns= -Dstorm.log.dir=/var/log/Bigdata/streaming/supervisor -Dlogging.sensitivity=S3 -Dlog4j.configurationFile=/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/etc/1_11_Supervisor/worker.xml -Dstorm.id=word-co
unt-8-1528079712 -Dworker.id=59d54d8c-f5a7-41b8-bf83-db54ba5eeab1 -Dworker.host=187-7-60-110 -Dworker.port=29100 -Dproc_backtype.storm.daemon.worker -cp /opt/huawei/Bigdata/FusionInsight_V100R002C60U20/Fusio
nInsight-Streaming-0.10.0/streaming/lib/xmlsec-1.5.7.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/log4j-api-2.5.jar:/opt/huawei/Bigdata/FusionInsight_V100
R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/log4j-over-slf4j-1.6.6.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/disruptor-2.10.4.jar:/opt/huaw
ei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/jul-to-slf4j-1.7.5.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/not-y
et-commons-ssl-0.3.9.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/bcprov-jdk15on-1.51.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-S
treaming-0.10.0/streaming/lib/joda-time-2.3.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/cas-client-core-hw-3.3.3.jar:/opt/huawei/Bigdata/FusionInsight_V1
00R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/xmltooling-1.4.5.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/log4j-slf4j-impl-2.5.jar:/opt/huaw
ei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/hadoop-auth-2.7.2.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/common
s-httpclient-3.1.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/minlog-1.2.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10
.0/streaming/lib/openws-1.5.5.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/reflectasm-1.07-shaded.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/Fus
ionInsight-Streaming-0.10.0/streaming/lib/opensaml-2.6.5.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/commons-codec-1.6.jar:/opt/huawei/Bigdata/FusionInsi
ght_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/clojure-1.6.0.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/asm-4.0.jar:/opt/huawei/Bigdata
/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/log4j-core-2.5.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/om-controller-api-0
.0.1.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/lib/kryo-2.21.jar:/opt/huawei/Bigdata/FusionInsight_V100R002C60U20/FusionInsight-Streaming-0.10.0/streaming/
lib/st
```

----End

15.18.9 Internal Server Error Is Displayed When the User Queries Information on the UI

Symptom

An MRS cluster is installed, and ZooKeeper and Storm are installed in the cluster.

"Internal Server Error" is displayed when a user accesses information from the **Storm Status** page of MRS Manager.

The detailed information is as follows:

```
Internal Server Error
org.apache.thrift7.transport.TTransportException: Frame size (306030) larger than max length (1048576)!
```

Possible Causes

- Nimbus of Storm is abnormal.
- Storm cluster information exceeds the default Thrift transmission size.

Cause Analysis

1. Check the Storm service status and monitoring metrics:
   - MRS Manager: Log in to MRS Manager and choose **Services** > **Storm**. Check the Storm status. The status is **Good**, and the monitoring metrics are correctly displayed.
   - FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > *Name of the target cluster* > **Service** > **Storm**.

     Check the Storm status. It is found that the status is good and the monitoring metrics are correctly displayed.
2. Click the **Instance** tab and check the status of the Nimbus instance. The status is normal.
3. Check the Thrift configuration of the Storm cluster. It is found that **nimbus.thrift.max_buffer_size** is set to **1048576** (1 MB).
4. The preceding configuration is the same as that in the exception information, indicating that the buffer size of Thrift is less than that required by the cluster information.

Procedure

Adjust the Thrift buffer size of the Storm cluster.

Step 1 Access the Storm parameter configuration page.

- MRS Manager: Log in to MRS Manager, choose **Services** > **Storm** > **Service Configuration**, and select **All** from the **Type** drop-down list.
- Operation on FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > **Services** > **Yarn** > **Configurations** > **All Configurations**.

Step 2 Change the value of **nimbus.thrift.max_buffer_size** to **10485760** (10 MB).

Step 3 Click Save Configuration and select **Restart the affected services or instances**. Click **OK** to restart the services.

----End

15.19 Using Ranger

15.19.1 After Ranger Authentication Is Enabled for Hive, Unauthorized Tables and Databases Can Be Viewed on the Hue Page

Issue

Although Ranger authentication is enabled for Hive, unauthorized tables and databases can be still viewed on the Hue page.

Symptom

In a normal cluster with Kerberos authentication disabled, after Ranger authentication is enabled for Hive, unauthorized tables and databases can be viewed on the Hue page.

Cause Analysis

After Ranger authentication is enabled for Hive, the default Hive policies contain two public group policies about databases. All users belong to the public group. By default, the public group is granted the permission to create tables in the default database and create other databases. Therefore, all users have the **show databases** and **show tables** permissions by default. If some users do not need to have these two permissions, you can delete the default public group policies on the Ranger web UI and grant the required user permissions.

Procedure

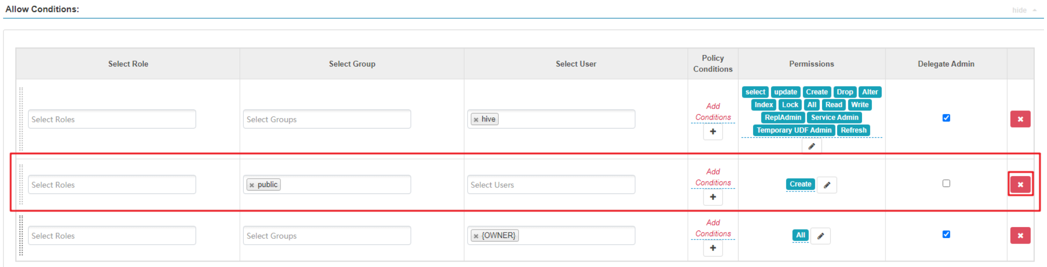
Step 1 Log in to the Ranger web UI.

Step 2 In the **Service Manager** area, click the Hive component name to access the Hive security access policy page.

Step 3 Click 📝 in the rows containing the **all - database** and **default database tables columns** policies.

Step 4 Delete the public group policies.

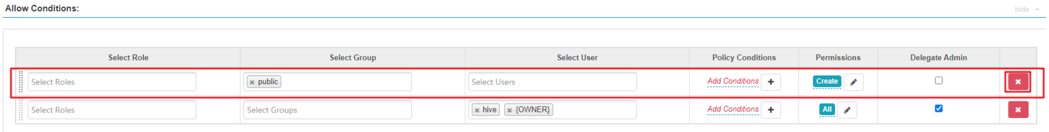
Figure 15-59 all - database policy



Allow Conditions:

Select Role	Select Group	Select User	Policy Conditions	Permissions	Delegate Admin	
Select Roles	Select Groups	× hive	Add Conditions +	select update Create Drop Alter Index Lock All Read Write ReplAdmin Service Admin Temporary UDF Admin Refresh	☑	✖
Select Roles	× public	Select Users	Add Conditions +	Create	□	✖
Select Roles	Select Groups	× {OWNER}	Add Conditions +	All	☑	✖

Figure 15-60 default database tables columns policy



Allow Conditions:

Select Role	Select Group	Select User	Policy Conditions	Permissions	Delegate Admin	
Select Roles	× public	Select Users	Add Conditions +	Create	□	✖
Select Roles	Select Groups	× hive × {OWNER}	Add Conditions +	All	☑	✖

Step 5 On the Hive security access policy page, click **Add New Policy** to add resource access policies for related users or user groups.

----End

15.20 Using Yarn

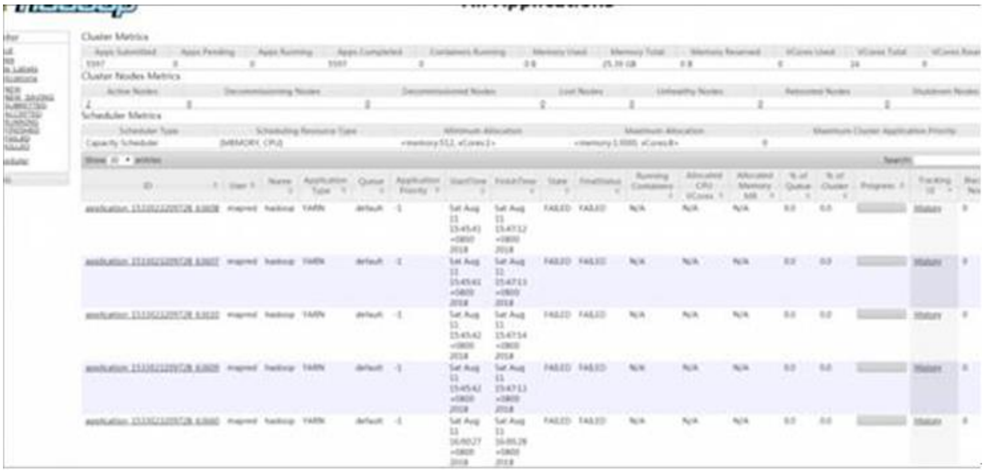
15.20.1 Plenty of Jobs Are Found After Yarn Is Started

Issue

After Yarn starts in an MRS cluster (MRS 2.x or earlier), plenty of jobs occupying resources are found.

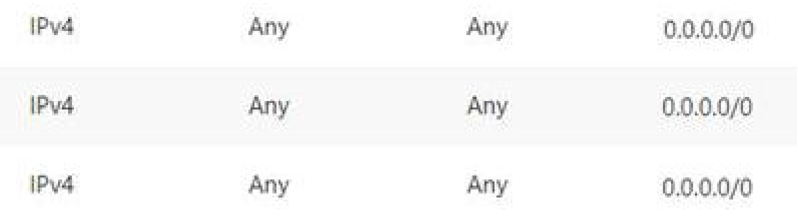
Symptom

After the customer creates an MRS cluster and starts Yarn, plenty of jobs occupying resources are found.



Cause Analysis

- It is suspected that there are hacker attacks.
- Set the Any protocol in the inbound direction of the SG to the 0.0.0.0/0.



IPv4	Any	Any	0.0.0.0/0
IPv4	Any	Any	0.0.0.0/0
IPv4	Any	Any	0.0.0.0/0

Procedure

Step 1 Log in to the MRS management console. On the **Active Clusters** page, click the cluster name. The cluster details page is displayed.

Step 2 Click **Manage** next to **Cluster Manager**. The **Access MRS Manager** page is displayed.

Step 3 Click **Manage Security Group Rule** to check the security group rule configuration.

Step 4 Check whether the source address of the Any protocol in the inbound direction is 0.0.0.0/0.

Step 5 If it is 0.0.0.0/0, change the remote end of the Any protocol in the inbound direction to a specified IP address. If it is not 0.0.0.0/0, there is no need to change the value.

Step 6 After the value is changed successfully, restart the cluster VM.

----End

Summary and Suggestions

Disable the Any protocol in the inbound direction, or specify the remote end of the Any protocol in the inbound direction as the specified IP address.

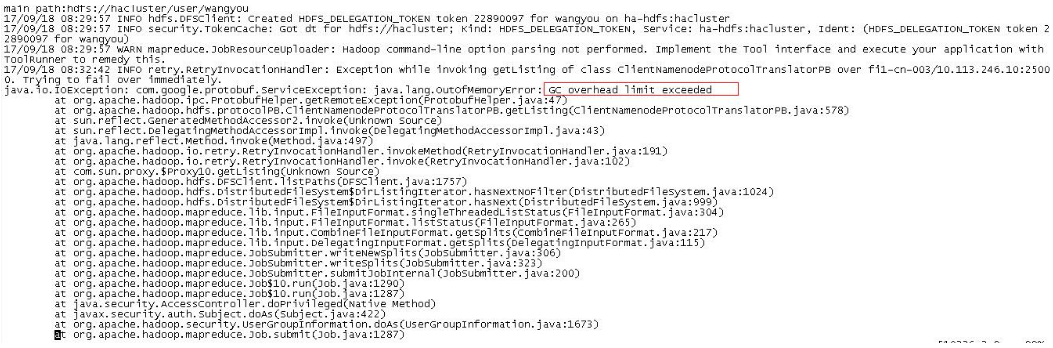
Related Information

For details, see **MapReduce Service User Guide** > **Security > Security Configuration Suggestions for Clusters with Kerberos Authentication Disabled**.

15.20.2 "GC overhead" Is Displayed on the Client When Tasks Are Submitted Using the Hadoop Jar Command

Symptom

When a user submits a task on the client, the client returns a memory overflow error.



```
main path:hdfs://hacluster/user/wangyou
17/09/18 08:29:57 INFO hdfs.DFSClient: Created HDFS_DELEGATION_TOKEN token 22890097 for wangyou on ha-hdfs:hacluster
17/09/18 08:29:57 INFO security.TokenCache: Got dt for hdfs://hacluster; Kind: HDFS_DELEGATION_TOKEN, Service: ha-hdfs:hacluster, Ident: (HDFS_DELEGATION_TOKEN token 2
2890097 for wangyou)
17/09/18 08:29:57 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with
ToolRunner to remedy this.
17/09/18 08:32:42 INFO retry.RetryInvocationHandler: Exception while invoking getListing of class ClientNamenodeProtocolTranslatorPB over fi1-cn-003/10.113.246.10:2500
0. Trying to fail over immediately.
java.io.IOException: com.google.protobuf.ServiceException: java.lang.OutOfMemoryError: GC overhead limit exceeded
        at org.apache.hadoop.ipc.ProtobufHelper.getRemoteException(ProtobufHelper.java:47)
        at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.getListing(ClientNamenodeProtocolTranslatorPB.java:578)
        at sun.reflect.GeneratedMethodAccessor2.invoke(Unknown Source)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:497)
        at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:191)
        at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:102)
        at com.sun.proxy.$Proxy10.getListing(Unknown Source)
        at org.apache.hadoop.hdfs.DFSClient.listPaths(DFSClient.java:1757)
        at org.apache.hadoop.hdfs.DistributedFileSystem$DirListingIterator.hasNextNoFilter(DistributedFileSystem.java:1024)
        at org.apache.hadoop.hdfs.DistributedFileSystem$DirListingIterator.hasNext(DistributedFileSystem.java:999)
        at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.singleThreadedListStatus(FileInputFormat.java:304)
        at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.listStatus(FileInputFormat.java:265)
        at org.apache.hadoop.mapreduce.lib.input.CombineFileInputFormat.getSplits(CombineFileInputFormat.java:217)
        at org.apache.hadoop.mapreduce.lib.input.DelegatingInputFormat.getSplits(DelegatingInputFormat.java:115)
        at org.apache.hadoop.mapreduce.JobSubmitter.writeNewSplits(JobSubmitter.java:306)
        at org.apache.hadoop.mapreduce.JobSubmitter.writeSplits(JobSubmitter.java:323)
        at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:200)
        at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1290)
        at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1287)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1673)
        at org.apache.hadoop.mapreduce.Job.submit(Job.java:1287)
```

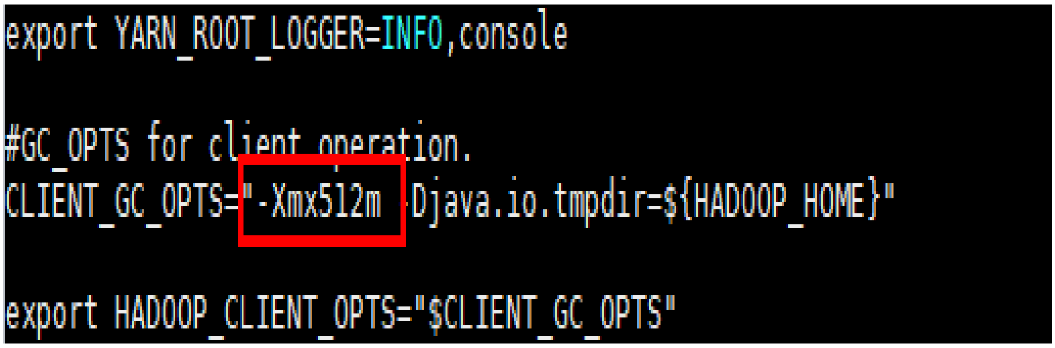
Cause Analysis

According to the error stack, the memory overflows when the HDFS files are read during task submission. Generally, the memory is insufficient because the task needs to read a large number of small files.

Solution

Step 1 Check whether multiple HDFS files need to be read for the started MapReduce tasks. If yes, reduce the file quantity by combining the small-sized files in advance or using **combineInputFormat**.

Step 2 Increase the memory when the **hadoop** command is run. The memory is set on the client. Change the value of **-Xmx** in **CLIENT_GC_OPTS** in the *Client installation directory***/HDFS/component_env** file to a larger value, for example, 512 MB. Run the **source component_env** command for the modification to take effect.



```
export YARN_ROOT_LOGGER=INFO,console

#GC_OPTS for client operation.
CLIENT_GC_OPTS="-Xmx512m -Djava.io.tmpdir=${HADOOP_HOME}"

export HADOOP_CLIENT_OPTS="$CLIENT_GC_OPTS"
```

----End

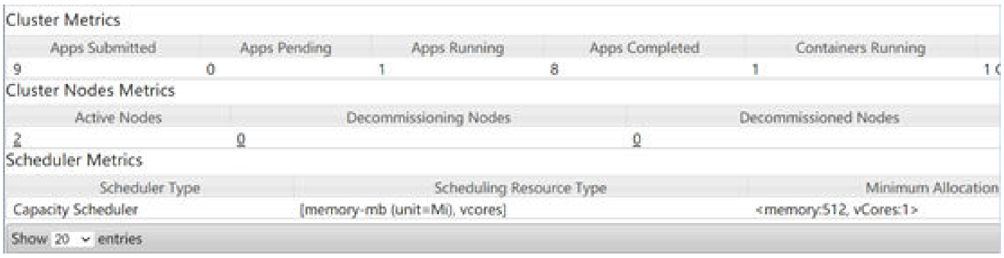
15.20.3 Disk Space Is Used Up Due to Oversized Aggregated Logs of Yarn

Issue

The disk usage of the cluster is high.

Symptom

- On the host management page of Manager, the disk usage is too high.
- Only a few tasks are running on the Yarn web UI.



Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	
9	0	1	8	1	1 (

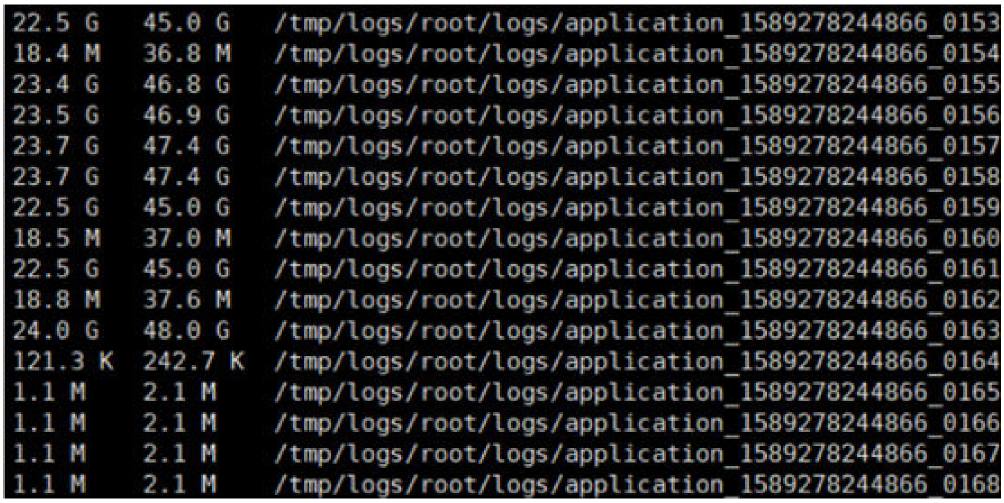
Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
2	0	0

Scheduler Metrics

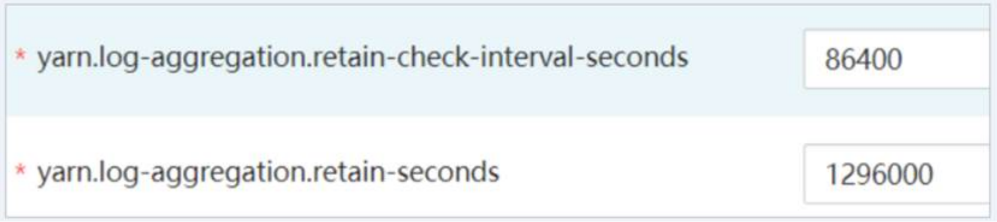
Scheduler Type	Scheduling Resource Type	Minimum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:512, vCores:1>

- After the **hdfs dfs -du -h /** command is executed on the master node of the cluster, the command output shows that the following files consume a large amount of disk space.



```
22.5 G   45.0 G   /tmp/logs/root/logs/application_1589278244866_0153
18.4 M   36.8 M   /tmp/logs/root/logs/application_1589278244866_0154
23.4 G   46.8 G   /tmp/logs/root/logs/application_1589278244866_0155
23.5 G   46.9 G   /tmp/logs/root/logs/application_1589278244866_0156
23.7 G   47.4 G   /tmp/logs/root/logs/application_1589278244866_0157
23.7 G   47.4 G   /tmp/logs/root/logs/application_1589278244866_0158
22.5 G   45.0 G   /tmp/logs/root/logs/application_1589278244866_0159
18.5 M   37.0 M   /tmp/logs/root/logs/application_1589278244866_0160
22.5 G   45.0 G   /tmp/logs/root/logs/application_1589278244866_0161
18.8 M   37.6 M   /tmp/logs/root/logs/application_1589278244866_0162
24.0 G   48.0 G   /tmp/logs/root/logs/application_1589278244866_0163
121.3 K  242.7 K  /tmp/logs/root/logs/application_1589278244866_0164
1.1 M    2.1 M    /tmp/logs/root/logs/application_1589278244866_0165
1.1 M    2.1 M    /tmp/logs/root/logs/application_1589278244866_0166
1.1 M    2.1 M    /tmp/logs/root/logs/application_1589278244866_0167
1.1 M    2.1 M    /tmp/logs/root/logs/application_1589278244866_0168
```

- The log aggregation configuration of the Yarn service is as follows.



* yarn.log-aggregation.retain-check-interval-seconds 86400

* yarn.log-aggregation.retain-seconds 1296000

Cause Analysis

Jobs are submitted too frequently, and the time for deleting aggregated log files is set to 1296000, that is, aggregated logs are retained for 15 days. As a result, aggregated logs cannot be released within a short period of time, exhausting the disk space.

Procedure

Step 1 Log in to Manager and navigate to the all configurations page of the MapReduce service.

- MRS Manager: Log in to MRS Manager, choose **Services** > **MapReduce** > **Service Configuration**, and select **All** from the **Type** drop-down list.
- FusionInsight Manager: Log in to FusionInsight Manager and choose **Cluster** > **Services** > **MapReduce**. On the MapReduce page, choose **Configurations** > **All Configurations**.

Step 2 Search for the **yarn.log-aggregation.retain-seconds** parameter and decrease its value based on site requirements, for example, to **259200**. In this case, the aggregated logs of Yarn are retained for three days, and the disk space is automatically released after the retention period expires.

Step 3 Click **Save Configuration** and deselect **Restart the affected services or instances**.

Step 4 Restart the MapReduce service during off-peak hours. The restart will interrupt upper-layer services and affect cluster management, maintenance, and services.

1. Log in to Manager.
2. Restart the MapReduce service.

----End

15.20.4 Temporary Files Are Not Deleted When an MR Job Is Abnormal

Issue

Temporary files are not deleted when an MR job is abnormal.

Symptom

There are too many files in the HDFS temporary directory, occupying too much memory.

Cause Analysis

When an MR job is submitted, related configuration files, JAR files, and files added by running the **-files** command are stored in the temporary directory on HDFS so that the started container can obtain the files. The configuration item **yarn.app.mapreduce.am.staging-dir** specifies the storage path. The default value is **/tmp/hadoop-yarn/staging**.

After a properly running MR job is complete, temporary files are deleted. However, when a Yarn task corresponding to the job exits abnormally, temporary files are not deleted. As a result, the number of files in the temporary directory increases over time, occupying more and more storage space.

Procedure

Step 1 Log in to a cluster.

1. Log in to any master node as user **root**. The user password is the one defined during cluster creation.
2. If Kerberos authentication is enabled for the cluster, run the following commands to go to the client installation directory and configure environment variables. Then, authenticate the user and enter the password as prompted. Obtain the password from an administrator.

   cd *Client installation directory*

   source bigdata_env

   kinit hdfs
3. If Kerberos authentication is not enabled for the cluster, run the following commands to switch to user **omm** and go to the client installation directory to configure environment variables:

   su - omm

   cd *Client installation directory*

   source bigdata_env

Step 2 Obtain the file list.

hdfs dfs -ls /tmp/hadoop-yarn/staging/*/.staging/ | grep "^drwx" | awk '{print $8}' > job_file_list

The **job_file_list** file contains the folder list of all jobs. The following shows an example of the file content:

```
/tmp/hadoop-yarn/staging/omm/.staging/job__<Timestamp>_<ID>
```

Step 3 Collect statistics on running jobs.

mapred job -list 2>/dev/null | grep job_ | awk '{print $1}' > run_job_list

The **run_job_list** file contains the IDs of running jobs. The content format is as follows:

```
job_<Timestamp>_<ID>
```

Step 4 Delete running jobs from the **job_file_list** file. Ensure that data of running jobs is not deleted by mistake when deleting expired data.

cat run_job_list | while read line; do sed -i "/$line/d" job_file_list; done

Step 5 Delete expired data.

cat job_file_list | while read line; do hdfs dfs -rm -r $line; done

Step 6 Delete temporary files.

rm -rf run_job_list job_file_list

----End

15.20.5 ResourceManager of Yarn (Port 8032) Throws Error "connection refused"

Issue

The ResourceManager of Yarn that requests to submit jobs throws error "connection refused", and the port number configured for Yarn is 8032.

Symptom

One of Yarn's ResourceManager nodes in the MRS cluster cannot be connected, and the port number configured for Yarn is 8032.

Cause Analysis

The service application runs outside the cluster, and the in-use client does not match the latest client configuration provided by the MRS cluster. The Yarn port is 8032, which is different from the actual port of Yarn's ResourceManager of MRS. As a result, the ResourceManager of Yarn that requests to submit jobs reports error "connection refused".

Procedure

Step 1 Update the MRS client.

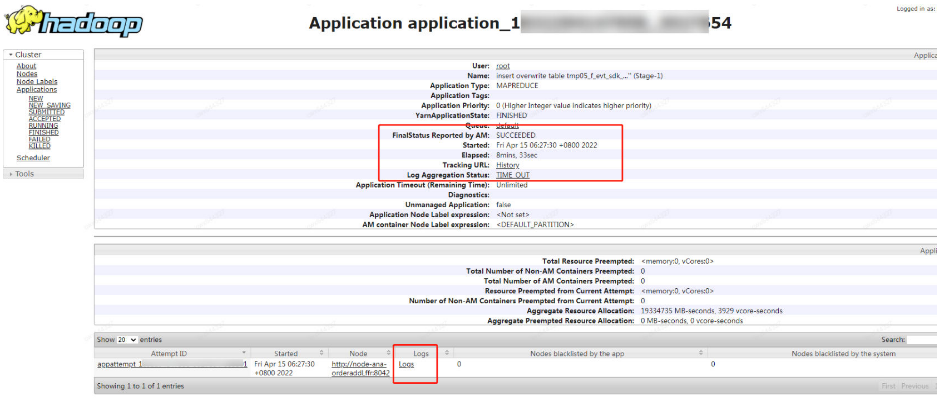
Step 2 Submit the job again.

----End

15.20.6 Failed to View Job Logs on the Yarn Web UI

Symptom

When a user logs in to the Yarn web UI to view job logs and clicks **Local logs**, error message "Could not access logs page!" is displayed.



Application application_1[redacted]54

Logged in as:

- Cluster
  - About
  - Nodes
  - Node Labels
  - Applications
    - NEW
    - NEW_SAVING
    - SUBMITTED
    - ACCEPTED
    - RUNNING
    - FINISHED
    - FAILED
    - KILLED
  - Scheduler
- Tools

Applica

User:	root
Name:	insert overwrite table tmp05_f_evt_sdk_..." (Stage-1)
Application Type:	MAPREDUCE
Application Tags:	
Application Priority:	0 (Higher Integer value indicates higher priority)
YarnApplicationState:	FINISHED
Queue:	default
FinalStatus Reported by AM:	SUCCEEDED
Started:	Fri Apr 15 06:27:30 +0800 2022
Elapsed:	8mins, 33sec
Tracking URL:	History
Log Aggregation Status:	TIME_OUT
Application Timeout (Remaining Time):	Unlimited
Diagnostics:	
Unmanaged Application:	false
Application Node Label expression:	<Not set>
AM container Node Label expression:	<DEFAULT_PARTITION>

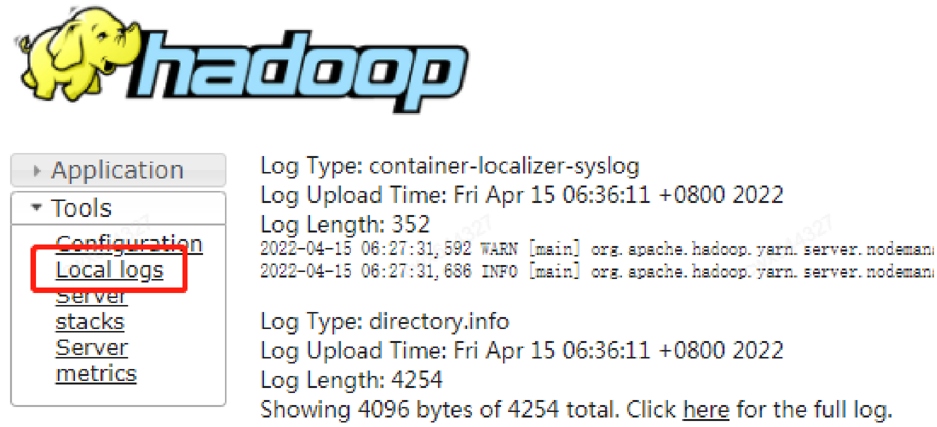
Appli

Total Resource Preempted:	<memory:0, vCores:0>
Total Number of Non-AM Containers Preempted:	0
Total Number of AM Containers Preempted:	0
Resource Preempted from Current Attempt:	<memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt:	0
Aggregate Resource Allocation:	19334735 MB-seconds, 3929 vcore-seconds
Aggregate Preempted Resource Allocation:	0 MB-seconds, 0 vcore-seconds

Show 20 entries Search:

Attempt ID	Started	Node	Logs	Nodes blacklisted by the app	Nodes blacklisted by the system
appattempt_1[redacted]1	Fri Apr 15 06:27:30 +0800 2022	http://node-ana-orderaddLffr:8042	Logs	0	0

Showing 1 to 1 of 1 entries First Previous



Log Type: container-localizer-syslog

Log Upload Time: Fri Apr 15 06:36:11 +0800 2022

Log Length: 352

```
2022-04-15 06:27:31,592 WARN [main] org.apache.hadoop.yarn.server.nodemana
2022-04-15 06:27:31,686 INFO [main] org.apache.hadoop.yarn.server.nodemana
```

Log Type: directory.info

Log Upload Time: Fri Apr 15 06:36:11 +0800 2022

Log Length: 4254

Showing 4096 bytes of 4254 total. Click here for the full log.

Cause Analysis

Local logs is used to access service logs. However, for security purposes, this function is inaccessible from the Yarn web UI. You can log in to the active ResourceManager node to view ResourceManager logs.

Procedure

Step 1 Log in to Manager and choose **Cluster** > **Services** > **Yarn**. On the **Yarn** page, click the **Instance** tab and take note of the service IP address of the active ResourceManager instance.

Step 2 Log in to the active ResourceManager node as user **root**.

Step 3 Go to the **/var/log/Bigdata/yarn/rm** directory and view the ResourceManager logs.

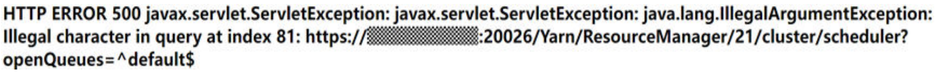
cd /var/log/Bigdata/yarn/rm

----End

15.20.7 An Error Is Reported When a Queue Name Is Clicked on the Yarn Page

Symptom

When Yarn uses the Capacity scheduler, error 500 is reported after a user clicks a queue name on the native Yarn web UI.



```
HTTP ERROR 500 javax.servlet.ServletException: javax.servlet.ServletException: java.lang.IllegalArgumentException:
Illegal character in query at index 81: https://[redacted]:20026/Yarn/ResourceManager/21/cluster/scheduler?
openQueues=^default$
```

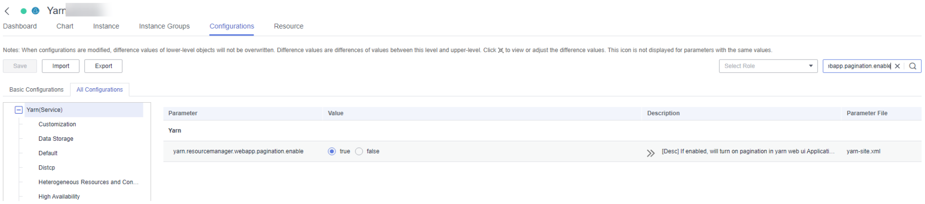
Cause Analysis

Symbol **^** in the URL cannot be identified. As a result, the page access fails.

Procedure

Step 1 Log in to Manager and choose **Cluster** > **Services** > **Yarn** > **Configurations** > **All Configurations**.

Step 2 Search for **yarn.resourcemanager.webapp.pagination.enable** in the search box.



Notes: When configurations are modified, difference values of lower-level objects will not be overwritten. Difference values are differences of values between this level and upper-level. Click ⌘ to view or adjust the difference values. This icon is not displayed for parameters with the same values.

ebapp.pagination.enable

Parameter	Value	Description	Parameter File
Yarn			
yarn.resourcemanager.webapp.pagination.enable	● true ○ false	[Desc] If enabled, will turn on pagination in yarn web ui Applicati...	yarn-site.xml

Step 3 If the value is **true** (default), change it to **false** and save the configuration.

Step 4 On the Yarn page, click **Instance**, select all ResourceManager instances, click **More**, and select **Instance Rolling Restart**. Wait until the instances are started.

----End

15.21 Using ZooKeeper

15.21.1 Accessing ZooKeeper from an MRS Cluster

Issue

An error is reported when a user attempts to access ZooKeeper from an MRS cluster.

Symptom

The customer uses **zkcli.sh** to access ZooKeeper on the MRS Master node, but an error is reported.

Cause Analysis

The command used by the customer is incorrect. As a result, an error is reported.

Procedure

Step 1 Obtain the ZooKeeper IP address.

Step 2 Log in to the Master node as user **root**.

Step 3 Run the following command to initialize environment variables:

source /opt/client/bigdata_env

Step 4 Run the **zkCli.sh -server** *IP address of the node where ZooKeeper is located*:**2181** command to connect to ZooKeeper of the MRS cluster.

The IP address of the node where ZooKeeper is located is the one queried in **Step 1**. Use commas (,) to separate multiple IP addresses.

Step 5 Run common commands such as **ls /** to view ZooKeeper information.

----End

15.22 Accessing OBS

15.22.1 When Using the MRS Multi-user Access to OBS Function, a User Does Not Have the Permission to Access the /tmp Directory

Issue

When the MRS multi-user access to OBS function is used to execute jobs such as Spark, Hive, and Presto jobs, an error message is displayed, indicating that the user does not have the permission to access the **/tmp** directory.

Symptom

When the MRS multi-user access to OBS function is used to execute jobs such as Spark, Hive, and Presto jobs, an error message is displayed, indicating that the user does not have the permission to access the **/tmp** directory.

Cause Analysis

A temporary directory exists during job execution. The user who submits the job does not have permission on the temporary directory.

Procedure

Step 1 On the **Dashboard** tab page of the cluster, query and record the name of the agency bound to the cluster.

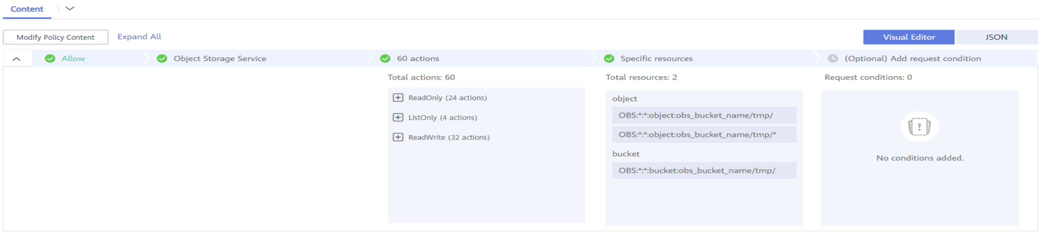
Step 2 Log in to the IAM console.

Step 3 Choose **Permissions**. On the displayed page, click **Create Custom Policy**.

- **Policy Name**: Enter a policy name.
- **Scope**: Select **Global services**.
- **Policy View**: Select **Visual editor**.
- **Policy Content**:
  a. **Allow**: Select **Allow**.
  b. **Select service**: Select **Object Storage Service (OBS)**.
  c. **Select action**: Select **WriteOnly**, **ReadOnly**, and **ListOnly**.
  d. **Specific resources**:
     i. Set **object** to **Specify resource path**, click **Add resource path**, and enter *obs_bucket_name*/**tmp/** and *obs_bucket_name*/**tmp/*** in **Path**. The **/tmp** directory is used as an example. If you need to add permissions for other directories, perform the following steps to add the directories and resource paths of all objects in the directories.
     ii. Set **bucket** to **Specify resource path**, click **Add resource path**, and enter *obs_bucket_name* in **Path**.

     Replace *obs_bucket-name* with the actual OBS bucket name. If the bucket type is Parallel File System, you need to add the *obs_bucket_name*/**tmp/** path. If the bucket type is Object Storage, you do not need to add the path.
  e. (Optional) Request condition, which does not need to be added currently.

Figure 15-61 Custom policy



Allow
Object Storage Service
60 actions
Specific resources
(Optional) Add request condition

Total actions: 60
ReadOnly (24 actions)
ListOnly (4 actions)
ReadWrite (32 actions)

Total resources: 2
object
OBS:*:*:object:obs_bucket_name/tmp/
OBS:*:*:object:obs_bucket_name/tmp/*
bucket
OBS:*:*:bucket:obs_bucket_name/tmp/

Request conditions: 0
No conditions added.

Step 4 Click **OK**.

Step 5 Select **Agency** and click **Assign Permissions** in the **Operation** column of the agency queried in **Step 1**.

Step 6 Query and select the created policy in **Step 3**.

Step 7 Click **OK**.

----End

15.22.2 When the Hadoop Client Is Used to Delete Data from OBS, It Does Not Have the Permission for the .Trash Directory

Issue

When a user uses the Hadoop client to delete data from OBS, an error message is displayed indicating that the user does not have the permission on the **.Trash** directory.

Symptom

After the **hadoop fs -rm obs://<***obs_path***>** command is executed, the following error information is displayed:

```
exception [java.nio.file.AccessDeniedException: user/root/.Trash/Current/: getFileStatus on user/root/.Trash/
Current/: status [403]
```

Cause Analysis

When deleting a file, Hadoop moves the file to the **.Trash** directory. If the user does not have the permission on the directory, error 403 is reported.

Procedure

Solution 1:

Run the **hadoop fs -rm -skipTrash** command to delete the file.

Solution 2:

Add the permission to access the **.Trash** directory to the agency corresponding to the cluster.

Step 1 On the **Dashboard** tab page of the cluster, query and record the name of the agency bound to the cluster.

Step 2 Log in to the IAM console.

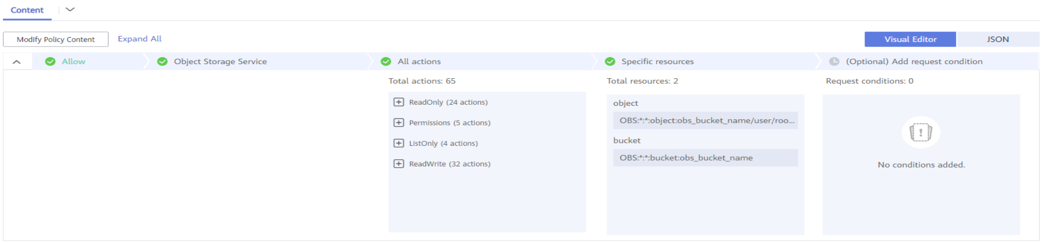
Step 3 Choose **Permissions**. On the displayed page, click **Create Custom Policy**.

- **Policy Name**: Enter a policy name.
- **Scope**: Select **Global services**.
- **Policy View**: Select **Visual editor**.
- **Policy Content**:
  a. **Allow**: Select **Allow**.
  b. **Select service**: Select **Object Storage Service (OBS)**.
  c. Select all operation permissions.
  d. **Specific resources**:
     i. Set **object** to **Specify resource path**, click **Add resource path**, and enter the **.Trash** directory, for example, **obs_bucket_name/user/root/.Trash/*** in **Path**.
     ii. Set **bucket** to **Specify resource path**, click **Add resource path**, and enter *obs_bucket_name* in **Path**.

Replace *obs_bucket-name* with the actual OBS bucket name.

e. (Optional) Request condition, which does not need to be added currently.

Figure 15-62 Custom policy



Allow | Object Storage Service | All actions | Specific resources | (Optional) Add request condition

Total actions: 65
- ReadOnly (24 actions)
- Permissions (5 actions)
- ListOnly (4 actions)
- ReadWrite (32 actions)

Total resources: 2

object
OBS:*:*:object:obs_bucket_name/user/roo...

bucket
OBS:*:*:bucket:obs_bucket_name

Request conditions: 0

No conditions added.

Step 4 Click **OK**.

Step 5 Select **Agency** and click **Assign Permissions** in the **Operation** column of the agency queried in **Step 1**.

Step 6 Query and select the created policy in **Step 3**.

Step 7 Click **OK**.

Step 8 Run the **hadoop fs -rm obs://<*obs_path*>** command again.

----End

16 Appendix

16.1 Precautions for MRS 3.x

Purpose

Custers of versions earlier than MRS 3.x use MRS Manager to manage and monitor MRS clusters. On the Cluster Management page of the MRS management console, you can view cluster details, manage nodes, components, alarms, patches, files, jobs, tenants, and backup and restoration. In addition, you can configure Bootstrap actions and manage tags.

MRS 3.x uses FusionInsight Manager to manage and monitor clusters. On the Cluster Management page of the MRS management console, you can view cluster details, manage nodes, components, alarms, files, jobs, Bootstrap actions, and tags.

Some maintenance operations of the MRS 3.x cluster are different from those of earlier versions. For details, see **MRS Manager Operation Guide (Applicable to 2.x and Earlier Versions)** and **FusionInsight Manager Operation Guide (Applicable to 3.x)**.

Accessing MRS Manager

- For details about how to access MRS Manager of versions earlier than MRS 3.x, see **Accessing MRS Manager MRS 2.x or Earlier)**.
- For details about how to access FusionInsight Manager of MRS 3.x, see **Accessing FusionInsight Manager (MRS 3.x or Later)**.

Modifying MRS Cluster Service Configuration Parameters

- For versions earlier than MRS 3.x, you can modify service configuration parameters on the cluster management page of the MRS management console.
  a. Log in to the MRS console. In the left navigation pane, choose **Clusters** > **Active Clusters**, and click a cluster name.
  b. Choose **Components** > *Name of the desired service* > **Service Configuration**.

The **Basic Configurations** tab page is displayed by default. To modify more parameters, click the **All Configurations** tab. The navigation tree displays all configuration parameters of the service. The level-1 nodes in the navigation tree are service names or role names. The parameter category is displayed after the level-1 node is expanded.

c. In the navigation tree, select the specified parameter category and change the parameter values on the right.

   If you are not sure about the location of a parameter, you can enter the parameter name in search box in the upper right corner. The system searches for the parameter in real time and displays the result.

d. Click **Save Configuration**. In the displayed dialog box, click **OK**.

e. Wait until the message "Operation succeeded" is displayed. Click **Finish**. The configuration is modified.

   Check whether there is any service whose configuration has expired in the cluster. If yes, restart the corresponding service or role instance for the configuration to take effect. You can also select **Restart the affected services or instances** when saving the configuration.

- In MRS 3.x, you need to log in to FusionInsight Manager to modify service configuration parameters.

  a. Log in to FusionInsight Manager.

  b. Choose **Cluster** > **Services**.

  c. Click the specified service name on the service management page.

  d. Click **Configurations**.

     The **Basic Configurations** tab page is displayed by default. To modify more parameters, click the **All Configurations** tab. The navigation tree displays all configuration parameters of the service. The level-1 nodes in the navigation tree are service names or role names. The parameter category is displayed after the level-1 node is expanded.

  e. In the navigation tree, select the specified parameter category and change the parameter values on the right.

     If you are not sure about the location of a parameter, you can enter the parameter name in search box in the upper right corner. The Manager searches for the parameter in real time and displays the result.

  f. Click **Save**. In the confirmation dialog box, click **OK**.

  g. Wait until the message "Operation succeeded" is displayed. Click **Finish**. The configuration is modified.

     Check whether there is any service whose configuration has expired in the cluster. If yes, restart the corresponding service or role instance for the configuration to take effect.

16.2 Installing the Flume Client

16.2.1 Installing the Flume Client on Clusters of Versions Earlier Than MRS 3.x

Scenario

To use Flume to collect logs, you must install the Flume client on a log host. You can create an ECS and install the Flume client on it.

This section applies to MRS 3.*x* or earlier clusters.

Prerequisites

- A streaming cluster with the Flume component has been created.
- The log host is in the same VPC and subnet with the MRS cluster.
- You have obtained the username and password for logging in to the log host.

Procedure

Step 1 Create an ECS that meets the requirements.

Step 2 Go to the cluster details page.

- For versions earlier than MRS 2.0.1, log in to MRS Manager and choose **Services**.
- For MRS 2.0.1 or later, click the cluster name on the MRS console and choose **Components**.

Step 3 Click **Download Client**.

1. In **Client Type**, select **All client files**.
2. In **Download to**, select **Remote host**.
3. Set **Host IP Address** to the IP address of the ECS, **Host Port** to **22**, and **Save Path** to .
   - If the default port **22** for logging in to an ECS through SSH has been changed, set **Host Port** to a new port.
   - The value of **Save Path** contains a maximum of 256 characters.
4. Set **Login User** to **root**.

   If another user is used, ensure that the user has permissions to read, write, and execute the save path.
5. Click **OK** to generate a client file.

   If the following information is displayed, the client package is saved.

   ```
   Client files downloaded to the remote host successfully.
   ```

   If the following information is displayed, check the username, password, and security group configurations of the remote host. Ensure that the username and password are correct and an inbound rule of the SSH (22) port has been added to the security group of the remote host. And then, go to **Step 3** to download the client again.

   ```
   Failed to connect to the server. Please check the network connection or parameter settings.
   ```

Step 4 Choose **Flume** > **Instance**. Query the **Business IP Address** of any Flume instance and any two MonitorServer instances.

Step 5 Log in to the ECS using VNC. See .

Log in to the ECS using an SSH key by referring to **Login Using an SSH Key** and set the password. Then log in to the ECS using VNC.

Step 6 On the ECS, switch to user **root** and copy the installation package to the **/opt** directory.

sudo su - root

cp /MRS_Flume_Client.tar /opt

Step 7 Run the following command in the **/opt** directory to decompress the package and obtain the verification file and the configuration package of the client:

tar -xvf MRS_Flume_Client.tar

Step 8 Run the following command to verify the configuration package of the client:

sha256sum -c MRS_Flume_ClientConfig.tar.sha256

If the following information is displayed, the file package is successfully verified:

```
MRS_Flume_ClientConfig.tar: OK
```

Step 9 Run the following command to decompress **MRS_Flume_ClientConfig.tar**:

tar -xvf MRS_Flume_ClientConfig.tar

Step 10 Run the following command to install the client running environment to a new directory, for example, **/opt/Flumeenv**. A directory is automatically generated during the client installation.

sh /opt/MRS_Flume_ClientConfig/install.sh /opt/Flumeenv

If the following information is displayed, the client running environment is successfully installed:

```
Components client installation is complete.
```

Step 11 Run the following command to configure environment variables:

source /opt/Flumeenv/bigdata_env

Step 12 Run the following commands to decompress the Flume client package:

cd /opt/MRS_Flume_ClientConfig/Flume

tar -xvf FusionInsight-Flume-1.6.0.tar.gz

Step 13 Run the following command to check whether the password of the current user has expired:

chage -l root

If the value of **Password expires** is earlier than the current time, the password has expired. Run the **chage -M -1 root** command to validate the password.

Step 14 Run the following command to install the Flume client to a new directory, for example, **/opt/FlumeClient**. A directory is automatically generated during the client installation.

sh /opt/MRS_Flume_ClientConfig/Flume/install.sh -d /opt/FlumeClient -f *service IP address of the MonitorServer instance* **-c** *path of the Flume configuration file* **-l /var/log/ -e** *service IP address of Flume* **-n** *name of the Flume client*

The parameters are described as follows:

- **-d**: indicates the installation path of the Flume client.
- (Optional) **-f**: indicates the service IP addresses of the two MonitorServer instances, separated by a comma (,). If the IP addresses are not configured, the Flume client will not send alarm information to MonitorServer, and the client information will not be displayed on MRS Manager.
- (Optional) **-c**: indicates the **properties.properties** configuration file that the Flume client loads after installation. If this parameter is not specified, the **fusioninsight-flume-1.6.0/conf/properties.properties** file in the client installation directory is used by default. The configuration file of the client is empty. You can modify the configuration file as required and the Flume client will load it automatically.
- (Optional) **-l**: indicates the log directory. The default value is **/var/log/Bigdata**.
- (Optional) **-e**: indicates the service IP address of the Flume instance. It is used to receive the monitoring indicators reported by the client.
- (Optional) **-n**: indicates the name of the Flume client.
- IBM JDK does not support **-Xloggc**. You must change **-Xloggc** to **-Xverbosegclog** in **flume/conf/flume-env.sh**. For 32-bit JDK, the value of **-Xmx** must not exceed 3.25 GB.
- In **flume/conf/flume-env.sh**, the default value of **-Xmx** is 4 GB. If the client memory is too small, you can change it to 512 MB or even 1 GB.

For example, run **sh install.sh -d /opt/FlumeClient**.

If the following information is displayed, the client is successfully installed:

```
install flume client successfully.
```

----End

16.2.2 Installing the Flume Client on MRS 3.*x* or Later Clusters

Scenario

To use Flume to collect logs, you must install the Flume client on a log host. You can create an ECS and install the Flume client on it.

This section applies to MRS 3.*x* or later clusters.

Prerequisites

- A cluster with the Flume component has been created.
- The log host is in the same VPC and subnet with the MRS cluster.
- You have obtained the username and password for logging in to the log host.

- The installation directory is automatically created if it does not exist. If it exists, the directory must be left blank. The directory path cannot contain any space.

Procedure

Step 1 Obtain the software package.

Log in to the FusionInsight Manager. Choose **Cluster** > *Name of the target cluster* > **Services** > **Flume**. On the Flume service page that is displayed, choose **More** > **Download Client** in the upper right corner and set **Select Client Type** to **Complete Client** to download the Flume service client file.

The file name of the client is **FusionInsight_Cluster_**<*Cluster ID*>**_Flume_Client.tar**. This section takes the client file **FusionInsight_Cluster_1_Flume_Client.tar** as an example.

Step 2 Upload the software package.

Upload the software package to a directory, for example, **/opt/client** on the node where the Flume service client will be installed as user **user**.

NOTE

user is the user who installs and runs the Flume client.

Step 3 Decompress the software package.

Log in to the node where the Flume service client is to be installed as user **user**. Go to the directory where the installation package is installed, for example, **/opt/client**, and run the following command to decompress the installation package to the current directory:

```
cd /opt/client
```

```
tar -xvf FusionInsight_Cluster_1_Flume_Client.tar
```

Step 4 Verify the software package.

Run the **sha256sum -c** command to verify the decompressed file. If **OK** is returned, the verification is successful. Example:

```
sha256sum -c FusionInsight_Cluster_1_Flume_ClientConfig.tar.sha256
```

```
FusionInsight_Cluster_1_Flume_ClientConfig.tar: OK
```

Step 5 Decompress the package.

```
tar -xvf FusionInsight_Cluster_1_Flume_ClientConfig.tar
```

Step 6 Run the following command in the Flume client installation directory to install the client to a specified directory (for example, **opt/FlumeClient**): After the client is installed successfully, the installation is complete.

```
cd /opt/client/FusionInsight_Cluster_1_Flume_ClientConfig/Flume/FlumeClient
```

./install.sh -d */opt/FlumeClient* **-f** *MonitorServerService IP address or host name of the role* **-c** *User service configuration filePath for storing properties.properties* **-s** *CPU threshold* **-l /var/log/Bigdata -e** *FlumeServer service IP address or host name* **-n** *Flume*

NOTE

- **-d**: Flume client installation path
- (Optional) **-f**: IP addresses or host names of two MonitorServer roles. The IP addresses or host names are separated by commas (,). If this parameter is not configured, the Flume client does not send alarm information to MonitorServer and information about the client cannot be viewed on the FusionInsight Manager GUI.
- (Optional) **-c**: Service configuration file, which needs to be generated on the configuration tool page of the Flume server based on your service requirements. Upload the file to any directory on the node where the client is to be installed. If this parameter is not specified during the installation, you can upload the generated service configuration file **properties.properties** to the **/opt/FlumeClient/fusioninsight-flume-1.9.0/conf** directory after the installation.
- (Optional) **-s**: cgroup threshold. The value is an integer ranging from 1 to 100 x *N*. *N* indicates the number of CPU cores. The default threshold is **-1**, indicating that the processes added to the cgroup are not restricted by the CPU usage.
- (Optional) **-l**: Log path. The default value is **/var/log/Bigdata**. The user **user** must have the write permission on the directory. When the client is installed for the first time, a subdirectory named **flume-client** is generated. After the installation, subdirectories named **flume-client-***n* will be generated in sequence. The letter *n* indicates a sequence number, which starts from 1 in ascending order. In the **/conf/** directory of the Flume client installation directory, modify the **ENV_VARS** file and search for the **FLUME_LOG_DIR** attribute to view the client log path.
- (Optional) **-e**: Service IP address or host name of FlumeServer, which is used to receive statistics for the monitoring indicator reported by the client.
- (Optional) **-n**: Name of the Flume client. You can choose **Cluster** > *Name of the desired cluster* > **Service** > **Flume** > **Flume Management** on FusionInsight Manager to view the client name on the corresponding node.
- If the following error message is displayed, run the **export JAVA_HOME=*JDK path*** command.

  ```
  JAVA_HOME is null in current user,please install the JDK and set the JAVA_HOME
  ```
- IBM JDK does not support **-Xloggc**. You must change **-Xloggc** to **-Xverbosegclog** in **flume/conf/flume-env.sh**. For 32-bit JDK, the value of **-Xmx** must not exceed 3.25 GB.
- When installing a cross-platform client in a cluster, go to the **/opt/client/FusionInsight_Cluster_1_Flume_ClientConfig/Flume/FusionInsight-Flume-1.9.0.tar.gz** directory to install the Flume client.

----End

17 Change History

Released On	What's New
2022-11-30	This issue is the seventh official release. ● The user account list adapts to 3.1.2-LTS.3. For details, see **User Account List**. ● Added section **Changing the Password for User compdbuser of the DBService Database**.
2022-07-30	This issue is the sixth official release. Released MRS 3.1.2-LTS.3. For details, see **Creating a Custom Cluster**.
2021-06-30	This issue is the fifth official release. Released MRS 3.1.0-LTS.1. For details, see **Creating a Custom Cluster**.
2020-10-24	This issue is the fourth official release. Added the following sections: ● **Methods of Creating MRS Clusters** ● **Quick Creation of a Hadoop Analysis Cluster** ● **Quick Creation of an HBase Analysis Cluster** **Quick Creation of a Kafka Streaming Cluster** Modified the following sections: ● **Creating a Custom Cluster** ● **Viewing Basic Cluster Information** ● **Viewing Information of a Historical Cluster**
2020-01-17	This issue is the third official release. Modified the following sections: ● **Creating a Custom Cluster** ● **Viewing Basic Cluster Information** ● **Viewing Information of a Historical Cluster**

Released On	What's New
2019-03-20	This issue is the second official issue. Added the following sections: - **Security Configuration Suggestions for Clusters with Kerberos Authentication Disabled** - **Authorizing O&M** - **Authorizing O&M** - **Adding a Tag to a Cluster** - **Installing Third-Party Software Using Bootstrap Actions** - **Restoring Patches for the Isolated Hosts** - **Rolling Restart** Modified the following sections: - **Creating a Custom Cluster** - **Viewing Basic Cluster Information** - **Configuring an Auto Scaling Rule** - **Viewing Information of a Historical Cluster**
2018-10-12	This issue is the first official release.