MapReduce Service

FAQs

Issue 01

Date 2023-01-11

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions


and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Contents

1 MRS Overview 1
1.1 What Is MRS Used For? 1
1.2 What Types of Distributed Storage Does MRS Support? 1
1.3 How Do I Create an MRS Cluster Using a Custom Security Group? 3
1.4 How Do I Use MRS? 3
1.5 Region and AZ 3
1.6 Can I Configure a Phoenix Connection Pool? 4
1.7 Does MRS Support Change of the Network Segment? 4
1.8 Can I Downgrade the Specifications of an MRS Cluster Node? 5
1.9 What Is the Relationship Between Hive and Other Components? 5
1.10 Does an MRS Cluster Support Hive on Spark? 5
1.11 What Are the Differences Between Hive Versions? 5
1.12 Which MRS Cluster Version Supports Hive Connection and User Synchronization? 6
1.13 What Are the Differences Between OBS and HDFS in Data Storage? 6
1.14 How Do I Obtain the Hadoop Pressure Test Tool? 6
1.15 What Is the Relationship Between Impala and Other Components? 6
1.16 Statement About the Public IP Addresses in the Open-Source Third-Party SDK Integrated by MRS 7
1.17 What Is the Relationship Between Kudu and HBase? 7
1.18 Does MRS Support Running Hive on Kudu? 7
1.19 What Are the Solutions for processing 1 Billion Data Records? 8
1.20 Can I Change the IP address of DBService? 8
1.21 Can I Clear MRS sudo Logs? 8
1.22 Is the Storm Log also limited to 20 GB in MRS cluster 2.1.0? 8
1.23 What Is Spark ThriftServer? 8
1.24 What Access Protocols Are Supported by Kafka? 9
1.25 What If Error 408 Is Reported When an MRS Node Accesses OBS? 9
1.26 What Is the Compression Ratio of zstd? 9
1.27 Why Are the HDFS, YARN, and MapReduce Components Unavailable When an MRS Cluster Is Bought? 9
1.28 Why Is the ZooKeeper Component Unavailable When an MRS Cluster Is Bought? 9
1.29 Which Python Versions Are Supported by Spark Tasks in an MRS 3.1.0 Cluster? 9
1.30 How Do I Enable Different Service Programs to Use Different YARN Queues? 10
1.31 Differences and Relationships Between the MRS Management Console and Cluster Manager 13

2 Billing 15
2.1 How Is MRS Billed? 15
2.2 Why Is the Price Not Displayed During MRS Cluster Creation? 15
2.3 How Is Auto Scaling Billed for an MRS Cluster? 15
2.4 How Is MRS Renewed? 16
2.5 How Is the Task Node in an MRS Cluster Billed? 16
2.6 Why Does My Unsubscription from ECS Fail After I Unsubscribe from MRS? 16

3 Account and Password 17
3.1 What Is the Account for Logging In to Manager? 17
3.2 How Do I Query and Change the Password Validity Period of an Account? 17

4 Accounts and Permissions 19
4.1 Does an MRS Cluster Support Access Permission Control If Kerberos Authentication Is not Enabled? 19
4.2 How Do I Assign Tenant Management Permission to a New Account? 19
4.3 How Do I Customize an MRS Policy? 20
4.4 Why Is the Manage User Function Unavailable on the System Page on MRS Manager? 20
4.5 Does Hue Support Account Permission Configuration? 21
4.6 Why Cannot I Submit Jobs on the Console After My IAM Account Is Assigned with Related Permissions? 21
4.7 How Do I Do If an Error Indicating Invalid Authentication Is Reported When I Submit an MRS Cluster Purchase Order? 21

5 Client Usage 23
5.1 How Do I Configure Environment Variables and Run Commands on a Component Client? 23
5.2 How Do I Disable ZooKeeper SASL Authentication? 23
5.3 An Error Is Reported When the kinit Command Is Executed on a Client Node Outside an MRS Cluster 23

6 Web Page Access 25
6.1 How Do I Change the Session Timeout Duration for an Open Source Component Web UI? 25
6.2 Why Cannot I Refresh the Dynamic Resource Plan Page on MRS Tenant Tab? 27
6.3 What Do I Do If the Kafka Topic Monitoring Tab Is Unavailable on Manager? 28
6.4 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? 28
6.5 How Do I Access HDFS of the Cluster in Security Mode on Windows Using EIPs? 29
6.6 How Do I Access HDFS of the Cluster in Normal Mode on Windows Using EIPs? 31
6.7 How Do I Access Hive of the Cluster in Security Mode on Windows Using EIPs? 34
6.8 How Do I Access Hive of the Cluster in Normal Mode on Windows Using EIPs? 36
6.9 How Do I Access Kafka of the Cluster in Security Mode on Windows Using EIPs? 38
6.10 How Do I Access Kafka of the Cluster in Normal Mode on Windows Using EIPs? 40
6.11 How Do I Access Spark of the Cluster in Security Mode on Windows Using EIPs? 41
6.12 How Do I Access Spark of the Cluster in Normal Mode on Windows Using EIPs? 44
6.13 How Do I Access HBase of the Cluster in Security Mode on Windows Using EIPs? 46
6.14 How Do I Access HBase of the Cluster in Normal Mode on Windows Using EIPs? 47

6.15 How Do I Switch the Mode of Accessing MRS Manager?....49

7 Alarm Monitoring....50

7.1 In an MRS Streaming Cluster, Can the Kafka Topic Monitoring Function Send Alarm Notifications?... 50

7.2 Where Can I View the Running Resource Queues When the Alarm "ALM-18022 Insufficient Yarn Queue Resources" Is Reported?.... 50

7.3 How Do I Understand the Multi-Level Chart Statistics in the HBase Operation Requests Metric?....50

8 Performance Tuning.... 52

8.1 Does an MRS Cluster Support System Reinstallation?.... 52

8.2 Can I Change the OS of an MRS Cluster?.... 52

8.3 How Do I Improve the Resource Utilization of Core Nodes in a Cluster?....52

8.4 How Do I Stop the Firewall Service?.... 53

9 Job Development....54

9.1 How Do I Get My Data into OBS or HDFS?.... 54

9.2 What Types of Spark Jobs Can Be Submitted in a Cluster?.... 55

9.3 Can I Run Multiple Spark Tasks at the Same Time After the Minimum Tenant Resources of an MRS Cluster Is Changed to 0?....55

9.4 How Do I Do If Job Parameters Separated By Spaces Cannot Be Identified?.... 55

9.5 What Are the Differences Between the Client Mode and Cluster Mode of Spark Jobs?....55

9.6 How Do I View MRS Job Logs?.... 56

9.7 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?....57

9.8 LauncherJob Job Execution Is Failed And the Error Message "jobPropertiesMap is null." Is Displayed 57

9.9 How Do I Do If the Flink Job Status on the MRS Console Is Inconsistent with That on Yarn?.... 57

9.10 How Do I Do If a SparkStreaming Job Fails After Being Executed Dozens of Hours and the OBS Access 403 Error is Reported?.... 57

9.11 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?.... 58

9.12 How Do I Do If Error Message "java.io.IOException: Connection reset by peer" Is Displayed During the Execution of a Spark Job?....58

9.13 How Do I Do If Error Message "requestId=4971883851071737250" Is Displayed When a Spark Job Accesses OBS?.... 59

9.14 How Do I Do If the Spark Job Error "UnknownScannerExeception" Is Reported?.... 59

9.15 Why DataArtsStudio Occasionally Fail to Schedule Spark Jobs and the Rescheduling also Fails?....59

9.16 How Do I Do If a Flink Job Fails to Execute and the Error Message "java.lang.NoSuchFieldError: SECURITY_SSL_ENCRYPT_ENABLED" Is Displayed?....60

9.17 Why Submitted Yarn Job Cannot Be Viewed on the Web UI?.... 60

9.18 How Do I Modify the HDFS NameSpace (fs.defaultFS) of an Existing Cluster?.... 60

9.19 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?.... 61

9.20 How Do I Do If the Error Message "slot request timeout" Is Displayed When I Submit a Flink Job? 61

9.21 Data Import and Export of DistCP Jobs.... 62

9.22 How Do I View SQL Statements for Hive Jobs on the YARN Web UI?..... 62

10 Cluster Upgrade/Patching..... 64

10.1 Can I Upgrade an MRS Cluster?..... 64

10.2 Can I Change the MRS Cluster Version?..... 64

11 Peripheral Ecosystem Interconnection..... 65

11.1 Can MRS Be Used to Perform Read and Write Operations on DLI Tables?..... 65

11.2 Does OBS Support the ListObjectsV2 Protocol?..... 65

11.3 Can MRS Data Be Stored in a Parallel File System Provided by OBS?..... 65

11.4 Can the Crawler Service Be Deployed in MRS?..... 65

11.5 Do DWS and MRS Support Secure Deletion (Preventing Retrieval After Deletion)?..... 65

11.6 Why Is the Kerberos-Authenticated MRS Cluster Not Found When a Connection Is Set Up from DLF?..... 66

11.7 How Do I Use PySpark on an ECS to Connect to an MRS Spark Cluster with Kerberos Authentication Enabled, on the Intranet?..... 66

11.8 Why Mapped Fields Do not Exist in the Database After HBase Synchronizes Data to CSS?..... 66

11.9 Can Flume Read Data from OBS?..... 66

11.10 Can MRS Connect to an External KDC?..... 66

11.11 How Do I Solve the Jetty Version Compatibility Issue in Open-Source Kylin 3.x and MRS 1.9.3 Interconnection?..... 66

11.12 What If Data Failed to Be Exported from MRS to an OBS Encrypted Bucket?..... 67

11.13 How Do I Install HSS on MRS Cluster Nodes?..... 67

12 Cluster Access..... 69

12.1 Can I Switch Between the Two Login Modes of MRS?..... 69

12.2 How Can I Obtain the IP Address and Port Number of a ZooKeeper Instance?..... 69

12.3 How Do I Access an MRS Cluster from a Node Outside the Cluster?..... 70

13 Big Data Service Development..... 72

13.1 Can MRS Run Multiple Flume Tasks at a Time?..... 72

13.2 How Do I Change FlumeClient Logs to Standard Logs?..... 72

13.3 Where Are the JAR Files and Environment Variables of Hadoop Stored?..... 73

13.4 What Compression Algorithms Does HBase Support?..... 73

13.5 Can MRS Write Data to HBase Through the HBase External Table of Hive?..... 73

13.6 How Do I View HBase Logs?..... 73

13.7 How Do I Set the TTL for an HBase Table?..... 73

13.8 How Do I Connect to HBase of MRS Through HappyBase?..... 74

13.9 How Do I Balance HDFS Data?..... 74

13.10 How Do I Change the Number of HDFS Replicas?..... 74

13.11 What Is the Port for Accessing HDFS Using Python?..... 75

13.12 How Do I Modify the HDFS Active/Standby Switchover Class?..... 79

13.13 What Is the Recommended Number Type of DynamoDB in Hive Tables?..... 79

13.14 Can the Hive Driver Be Interconnected with DBCP2?..... 79

13.15 How Do I View the Hive Table Created by Another User?..... 80

13.16 Where Can I Download the Dependency Package (com.huawei.gaussc10) in the Hive Sample Project? 81
13.17 Can I Export the Query Result of Hive Data? 82
13.18 How Do I Do If an Error Occurs When Hive Runs the **beeline -e** Command to Execute Multiple Statements? 82
13.19 How Do I Do If a "hivesql/hivescript" Job Fails to Submit After Hive Is Added? 82
13.20 What If an Excel File Downloaded on Hue Failed to Open? 83
13.21 How Do I Do If Sessions Are Not Released After Hue Connects to HiveServer and the Error Message "over max user connections" Is Displayed? 84
13.22 How Do I Reset Kafka Data? 85
13.23 How Do I Obtain the Client Version of MRS Kafka? 85
13.24 What Access Protocols Are Supported by Kafka? 85
13.25 How Do I Do If Error Message "Not Authorized to access group xxx" Is Displayed When a Kafka Topic Is Consumed? 85
13.26 What Compression Algorithms Does Kudu Support? 86
13.27 How Do I View Kudu Logs? 86
13.28 How Do I Handle the Kudu Service Exceptions Generated During Cluster Creation? 86
13.29 What Are the Differences Between Sample Project Building and Application Development? Is Python Code Supported? 87
13.30 Does OpenTSDB Support Python APIs? 87
13.31 How Do I Configure Other Data Sources on Presto? 87
13.32 How Do I Update the Ranger Certificate? 89
13.33 How Do I Connect to Spark Shell from MRS? 91
13.34 How Do I Connect to Spark Beeline from MRS? 91
13.35 Where Are the Execution Logs of Spark Jobs Stored? 92
13.36 How Do I Specify a Log Path When Submitting a Task in an MRS Storm Cluster? 92
13.37 How Do I Check Whether the ResourceManager Configuration of Yarn Is Correct? 93
13.38 How Do I Modify the allow_drop_detached Parameter of ClickHouse? 95
13.39 How Do I Do If an Alarm Indicating Insufficient Memory Is Reported During Spark Task Execution? 96
13.40 How Do I Do If ClickHouse Consumes Excessive CPU Resources? 96
13.41 How Do I Obtain a Spark JAR File? 96
13.42 Why Is an Alarm Generated When the NameNode Process Is Not Restarted After the hdfs-site.xml File Is Modified? 97
13.43 It Takes a Long Time for Spark SQL to Access Hive Partitioned Tables Before Job Startup 97

14 API 99

14.1 How Do I Configure the node_id Parameter When Using the API for Adjusting Cluster Nodes? 99

15 Cluster Management 100

15.1 How Do I View All Clusters? 100
15.2 How Do I View Log Information? 100
15.3 How Do I View Cluster Configuration Information? 101
15.4 How Do I Add Services to an MRS Cluster? 101
15.5 How Do I Install Kafka and Flume in an MRS Cluster? 101

15.6 How Do I Stop an MRS Cluster?.....101
15.7 Do I Need to Shut Down a Master Node Before Upgrading Its Specifications?.....102
15.8 Can I Expand Data Disk Capacity for MRS?.....102
15.9 Can I Add Components to an Existing Cluster?.....102
15.10 Can I Delete Components Installed in an MRS Cluster?.....102
15.11 Can I Change MRS Cluster Nodes on the MRS Console?.....102
15.12 How Do I Shield Cluster Alarm/Event Notifications?.....103
15.13 Why Is the Resource Pool Memory Displayed in the MRS Cluster Smaller Than the Actual Cluster Memory?.....103
15.14 How Do I Configure the knox Memory?.....103
15.15 What Is the Python Version Installed for an MRS Cluster?.....104
15.16 How Do I View the Configuration File Directory of Each Component?.....104
15.17 How Do I Upload a Local File to a Node Inside a Cluster?.....105
15.18 How Do I Do If the Time on MRS Nodes Is Incorrect?.....105
15.19 How Do I Query the Startup Time of an MRS Node?.....106
15.20 How Do I Do If Trust Relationships Between Nodes Are Abnormal?.....106
15.21 How Do I Adjust the Memory Size of the manager-executor Process?.....108
15.22 Can I Modify a Master Node in an Existing MRS Cluster?.....108

16 Kerberos Usage.....110
16.1 How Do I Change the Kerberos Authentication Status of a Created MRS Cluster?.....110
16.2 What Are the Ports of the Kerberos Authentication Service?.....110
16.3 How Do I Deploy the Kerberos Service in a Running Cluster?.....110
16.4 How Do I Access Hive in a Cluster with Kerberos Authentication Enabled?.....110
16.5 How Do I Access Presto in a Cluster with Kerberos Authentication Enabled?.....111
16.6 How Do I Access Spark in a Cluster with Kerberos Authentication Enabled?.....112
16.7 How Do I Prevent Kerberos Authentication Expiration?.....113

17 Metadata Management.....115
17.1 Where Can I View Hive Metadata?.....115

1 MRS Overview

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.

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 1-1** lists the component versions supported by MRS.

For details, see **List of MRS Component Versions**.

Table 1-1 MRS component versions

Component	MRS 1.9.2 (Applicable to MRS 1.9.*x*)	MRS 3.1.0
Alluxio	2.0.1	N/A
CarbonData	1.6.1	2.0.1
ClickHouse	N/A	21.3.4.25
DBService	1.0.0	2.7.0
Flink	1.7.0	1.12.0
Flume	1.6.0	1.9.0

Component	MRS 1.9.2 (Applicable to MRS 1.9.*x*)	MRS 3.1.0
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
MRS Manager	1.9.2	N/A
FusionInsight Manager	N/A	8.1.0

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.

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 Huawei 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. Develop a data processing program. For details about how to quickly develop such a program and execute it properly, see the sample code and tutorials provided in **Method of Building an MRS Sample Project**.
2. Upload local programs and data files to OBS.
3. 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.
4. Use MRS to submit, execute, and monitor your programs.
5. 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.
6. Terminate any MRS cluster that you do not require after job execution is complete. The terminated cluster is no longer billed.

1.5 Region and AZ

What Are Regions and AZs?

A region and availability zone (AZ) identify the location of a data center. You can create resources in regions and AZs.

shows the relationship between regions and AZs.

What Are the Considerations for Selecting a Region?

When selecting a region, consider the following factors:

- Location

  Select a region close to you or your target users to minimize network latency and increase access speed. However, there is no distinct difference between the infrastructure, BGP network quality, and resource operations and configurations in Chinese mainland regions. Therefore, you do not need to consider the network latency when selecting a region in Chinese mainland.

  - If you or your target users are in the Asia Pacific area excluding the Chinese mainland, select the **CN-Hong Kong**, **AP-Bangkok**, or **AP-Singapore** region.
  - If you or your target users are in Africa, select the **AF-Johannesburg** region.
  - If you or your target users are in Europe, select the **EU-Paris** region.
- Resource price

  Resource prices may vary in different regions. For details, see .

What Are the Considerations for Selecting an AZ?

Consider the requirements for disaster recovery (DR) and network latency when selecting an AZ:

- Deploy resources in different AZs in the same region for DR purposes.
- Deploy resources in the same AZ for minimum latency.

How Do I Obtain an Endpoint?

When you use resources by using an API, you must specify its regional endpoint. For details about Huawei Cloud regions and endpoints, see .

Not supported currently. Contact the enterprise administrator.

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. For details, see **https://stackoverflow.com/questions/35183713/what-is-the-correct-way-to-pool-the-phoenix-query-server-connections**.

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.

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.

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.

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.

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**.

1.12 Which MRS Cluster Version Supports Hive Connection and User Synchronization?

MRS 2.0.5 or later supports Hive connections on DataArts Studio and provides the IAM user synchronization function.

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.

1.14 How Do I Obtain the Hadoop Pressure Test Tool?

Download it from **https://github.com/Intel-bigdata/HiBench**.

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.

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.

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.

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.

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.

1.20 Can I Change the IP address of DBService?

MRS does not support the change of the DBService IP address.

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.

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.

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.

1.24 What Access Protocols Are Supported by Kafka?

Kafka supports PLAINTEXT, SSL, SASL_PLAINTEXT, and SASL_SSL.

1.25 What If Error 408 Is Reported When an MRS Node Accesses OBS?

Change the OBS domain name to a domain name ended with "myhuaweicloud.com".

1.26 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.

1.27 Why Are the HDFS, YARN, and MapReduce Components Unavailable When an MRS Cluster Is Bought?

The HDFS, YARN, and MapReduce components are integrated in Hadoop. If the three components are unavailable when are MRS cluster is bought, select Hadoop instead. After an MRS cluster is created, HDFS, YARN, and MapReduce are available in the **Components** page.

1.28 Why Is the ZooKeeper Component Unavailable When an MRS Cluster Is Bought?

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.

1.29 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.

1.30 How Do I Enable Different Service Programs to Use Different YARN Queues?

Create a tenant on Manager.

For details, see **Adding a Sub-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 1-2**.

Table 1-2 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.

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

1.31 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

Common Operation	MRS Console	FusionInsight Manager
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
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

2 Billing

2.1 How Is MRS Billed?

MRS supports two billing modes: Yearly or monthly subscription and pay-per-use.

- Yearly or monthly subscription: You pay for the cluster by year or month, in advance. The minimum usage duration is one month, and the maximum usage duration is one year.
- Pay-per-use: a postpaid billing mode. Nodes are charged by actual duration of use, with a billing cycle of one hour.

NOTE

- Clusters in the **Starting**, **Failed**, or **Terminated** state will not be charged.
- When you purchase an MRS cluster, the price displayed is only for the cost of the cluster. The usage for data storage, bandwidth, and traffic on MRS are billed separately.

2.2 Why Is the Price Not Displayed During MRS Cluster Creation?

When you are purchasing an MRS cluster, the price displayed at the bottom of the page will not include disk cost if you specify only the number of disks without configuring the instance count. The price will include the disk cost after you configure both the number of disks and instance count.

2.3 How Is Auto Scaling Billed for an MRS Cluster?

The price displayed at the bottom when you purchase an MRS cluster will not include the auto scaling fee if you specify only the auto scaling range for the Task nodes without configuring the instance count. The price will include the auto scaling fee after you configure both the auto scaling range and instance count.

If you add nodes through the auto scaling function, the added nodes will be billed by the actual usage duration per hour regardless of whether the cluster's billing mode is yearly/monthly or pay-per-use mode.

2.4 How Is MRS Renewed?

MRS provides two billing modes: pay-per-use and yearly/monthly subscription. In pay-per-use mode, billing is per hour and an insufficient balance may cause overdue payments. For yearly/monthly subscription, you need to renew your resources before they expire. Otherwise, your resources will be given a retention period. During this period, data will be retained, but MRS clusters will be stopped.

2.5 How Is the Task Node in an MRS Cluster Billed?

Task nodes in a cluster are billed in pay-per-use mode, regardless of the cluster billing mode. That is, they are billed per hour based on the actual usage duration.

2.6 Why Does My Unsubscription from ECS Fail After I Unsubscribe from MRS?

1. Check that the ECS ID is not used by an MRS cluster in use.
2. On the ECS console, find the ECS to be unsubscribed from and click **Locked by MRS** to unlock it.
3. Click **Unsubscribe** again.
4. If the unsubscription still fails, collect the ECS ID and contact Huawei Cloud technical support.

3 Account and Password

3.1 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.

3.2 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.

  The operations are as follows:

  Versions earlier than MRS 3.*x*: Change the password validity period of a human-machine user by referring to **Modifying a Password Policy**.

  MRS 3.*x* or later: Change the password validity period of a human-machine user by referring to**Configuring Password Policies**.

4 Accounts and Permissions

4.1 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.

For details about how to configure permissions, see **Permission Management**.

For details about how to use clusters with Kerberos authentication enabled, see **Clusters with Kerberos Authentication Enabled**.

4.2 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

4.3 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.

4.4 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.

4.5 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.

4.6 Why Cannot I Submit Jobs on the Console After My IAM Account Is Assigned with Related Permissions?

Symptom

I cannot submit jobs on the MRS console after my IAM account is assigned with the **MRS ReadOnlyAccess** and **MRS FullAccess** permissions.

Troubleshooting

The account has the **MRS ReadOnlyAccess** and **MRS FullAccess** permissions. Due to the permission priority, the account cannot add jobs on the MRS console.

The group to which the IAM account belongs has the **MRS FullAccess**, **MRS ReadOnlyAccess**, and **MRS Administrator** permissions. **MRS FullAccess** and **MRS ReadOnlyAccess** are fine-grained permissions, and **MRS Administrator** is an role-based access control (RBAC) policy. Fine-grained permissions have higher priorities over RBAC policies. If fine-grained permissions and RBAC policies are configured, fine-grained permissions take effect first. Fine-grained permissions follow the deny priority principle. Therefore, the **MRS ReadOnlyAccess** permission takes effect finally. So, a message is displayed, indicating that the account does not have the permission.

Delete the **MRS ReadOnlyAccess** permission, log out of the console, and log in to the console again.

4.7 How Do I Do If an Error Indicating Invalid Authentication Is Reported When I Submit an MRS Cluster Purchase Order?

Symptom

When I submit an order for buying an MRS cluster, an error message is displayed, indicating that the authentication is invalid. After I press **F12** to view the network request, error code 401 is displayed.

Troubleshooting

1. Check the backend API request logs. The alarm information is as follows: IAM users have not been assigned the **mrs:cluster:create** permission in the fine-grained policy.
2. IAM users on the customer side belong to multiple user groups, and different default MRS policies are assigned to these user groups. Low-permission policies are preferentially matched. This policy does not contain the **mrs:cluster:create** permission. As a result, the cluster creation operation cannot be submitted, and error 401 is reported.
3. After the user is removed from the user group corresponding to the low-permission policy, the cluster purchase order can be submitted.

5 Client Usage

5.1 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** *Client installation directory* 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.

5.2 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.

5.3 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.

6 Web Page Access

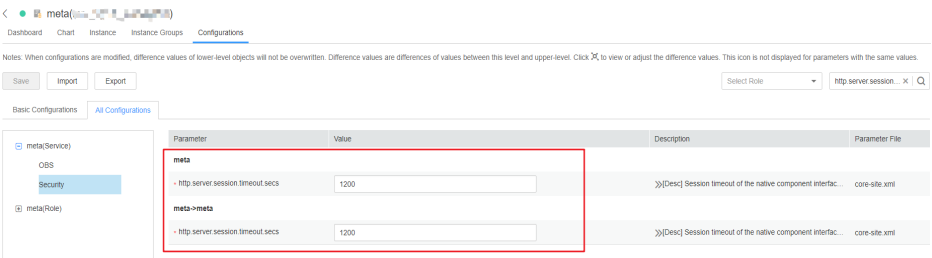
6.1 How Do I Change the Session Timeout Duration for an Open Source Component Web UI?

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.



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.

6.2 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

6.3 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

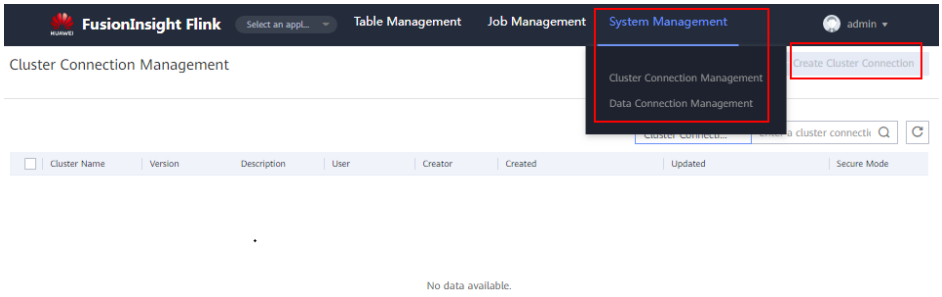
sh /opt/Bigdata/apache-tomcat-7.0.78/bin/startup.sh

----End

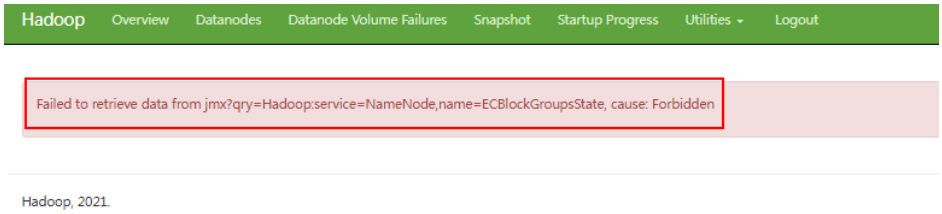
6.4 How Do I Do If an Error Is Reported or Some Functions Are Unavailable When I Access the Web UIs of HDFS, Hue, YARN, and Flink?

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:

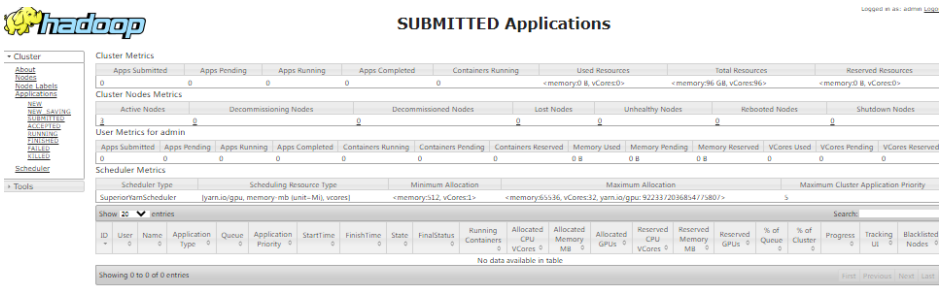
- After you log in to the web UI of Flink as the current user, some content cannot be displayed, and you do not have the permission to create applications, cluster connections, or data connections.



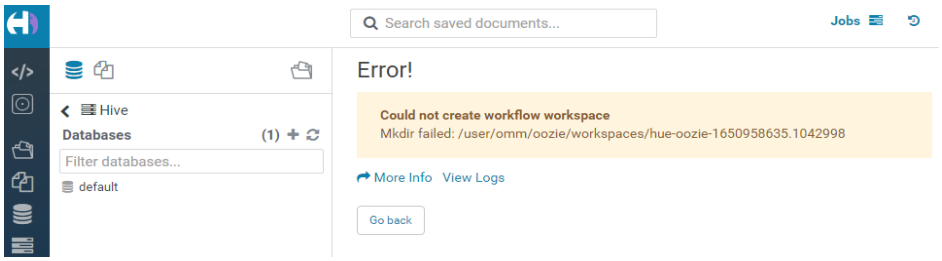
- After you log in to the web UI of HDFS as the current user, error message "Failed to retrieve data from /jmx?qry=java.lang:type=Memory, cause: Forbidden" is displayed.



- After you log in to the web UI of YARN as the current user, you cannot view job information.



- 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.



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. For details, see **Creating a User**.

6.5 How Do I Access HDFS of the Cluster in Security Mode on Windows Using EIPs?

Scenario

This section describes how to bind Elastic IP addresses (EIPs) to a cluster and configure HDFS files so that sample files can be compiled locally.

This section uses HdfsExample as an example.

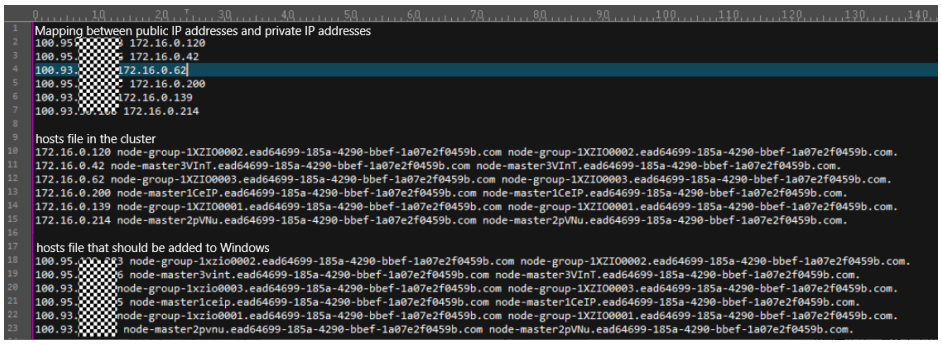
Procedure

Step 1 Apply for an EIP for each node in the cluster and add public IP addresses and corresponding host domain names of all nodes to the Windows local **hosts** file. (If a host name contains uppercase letters, change them to lowercase letters.)

1. On the VPC console, apply for EIPs (the number of EIPs you buy should be equal to the number of nodes in the cluster), click the name of each node in the MRS cluster, and bind an EIP to each node on the **EIPs** page.

   For details, see **Virtual Private Cloud** > **User Guide** > **EIP** > **Assigning an EIP and Binding It to an ECS**.

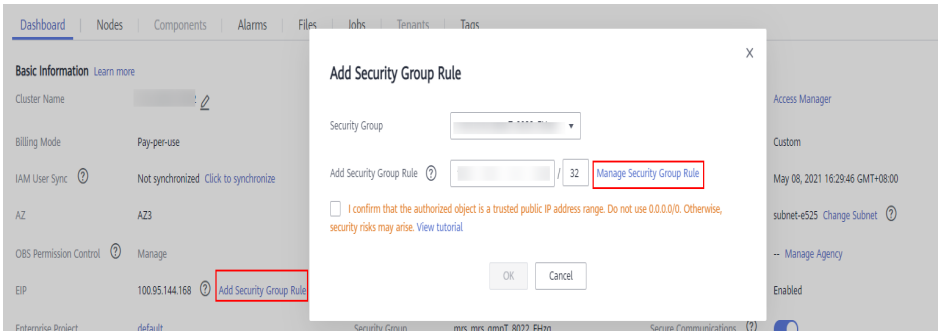
2. Record the mapping between the public IP addresses and private IP addresses. Change the private IP addresses in the **hosts** file to the corresponding public IP addresses.



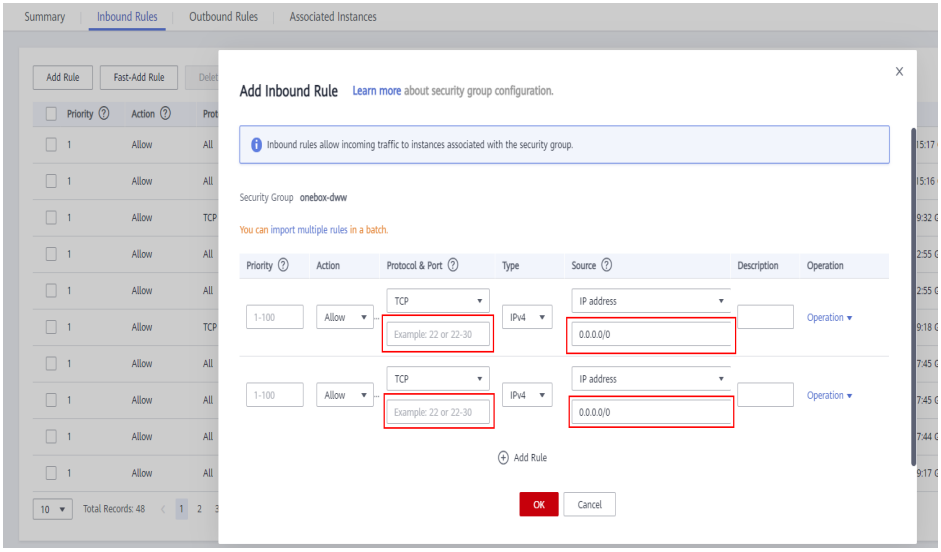
Step 2 Change the IP addresses in the **krb5.conf** file to the corresponding host names.

Step 3 Configure security group rules for the cluster.

1. On the **Dashboard** page, choose **Add Security Group Rule** > **Manage Security Group Rule**.



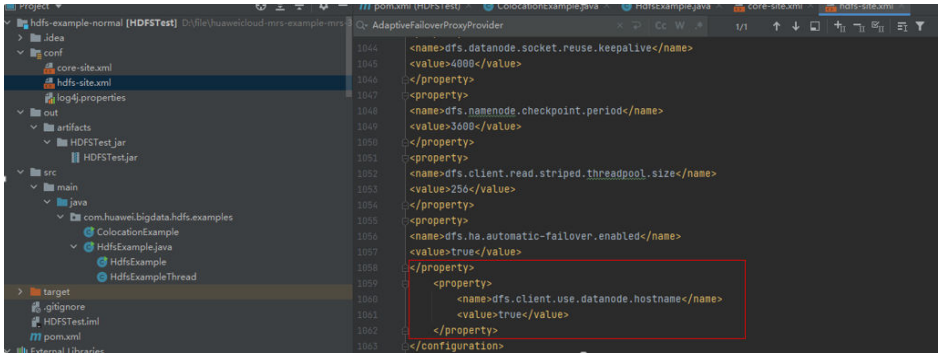
2. On the **Inbound Rules** tab page, click **Add Rule**. In the **Add Inbound Rule** dialog box, configure Windows IP addresses and ports 21730TCP, 21731TCP/UDP, and 21732TCP/UDP.



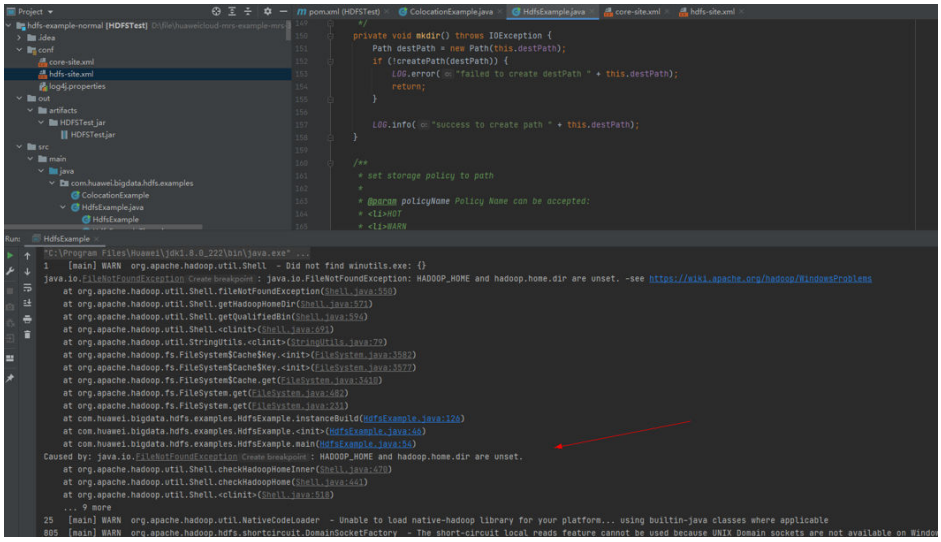
Step 4 On Manager, choose **Cluster** > **Services** > **HDFS** > **More** > **Download Client**, copy the **core-site.xml** and **hdfs-site.xml** files on the client to the **conf** directory of the sample project, and add the following content to the **hdfs-site.xml** file.

```
<property>
 <name>dfs.client.use.datanode.hostname</name>
 <value>true</value>
 </property>
```

(Change the DataNode communication mode to hostname.)



After the modification, an error message indicating that hadoop_home does not exist may be displayed when you run the sample project. You can ignore the error because it does not affect the use.



Step 5 Before running the sample code, change the value of **PRNCIPAL_NAME** in the sample code to the username for security authentication.

----End

6.6 How Do I Access HDFS of the Cluster in Normal Mode on Windows Using EIPs?

Scenario

This section describes how to bind Elastic IP addresses (EIPs) to a cluster and configure HDFS files so that sample files can be compiled locally.

This section uses HdfsExample as an example.

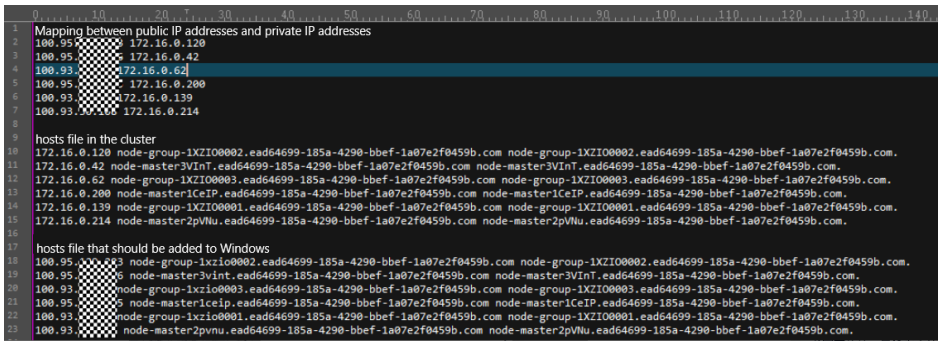
Procedure

Step 1 Apply for an EIP for each node in the cluster and add public IP addresses and corresponding host domain names of all nodes to the Windows local **hosts** file. (If a host name contains uppercase letters, change them to lowercase letters.)

1. On the VPC console, apply for EIPs (the number of EIPs you buy should be equal to the number of nodes in the cluster), click the name of each node in the MRS cluster, and bind an EIP to each node on the **EIPs** page.

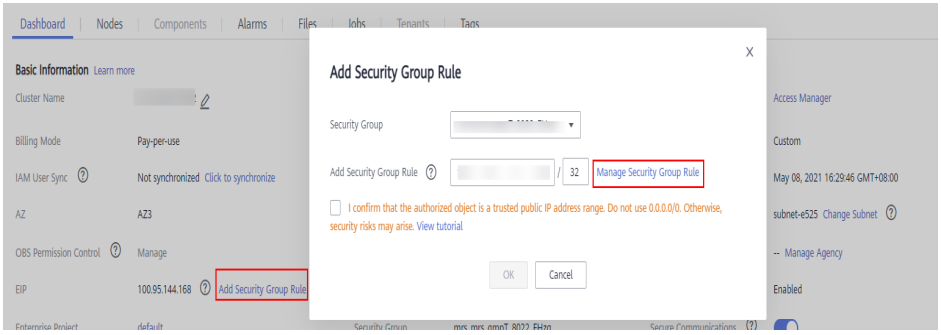
   For details, see **Virtual Private Cloud** > **User Guide** > **EIP** > **Assigning an EIP and Binding It to an ECS**.

2. Record the mapping between the public IP addresses and private IP addresses. Change the private IP addresses in the **hosts** file to the corresponding public IP addresses.

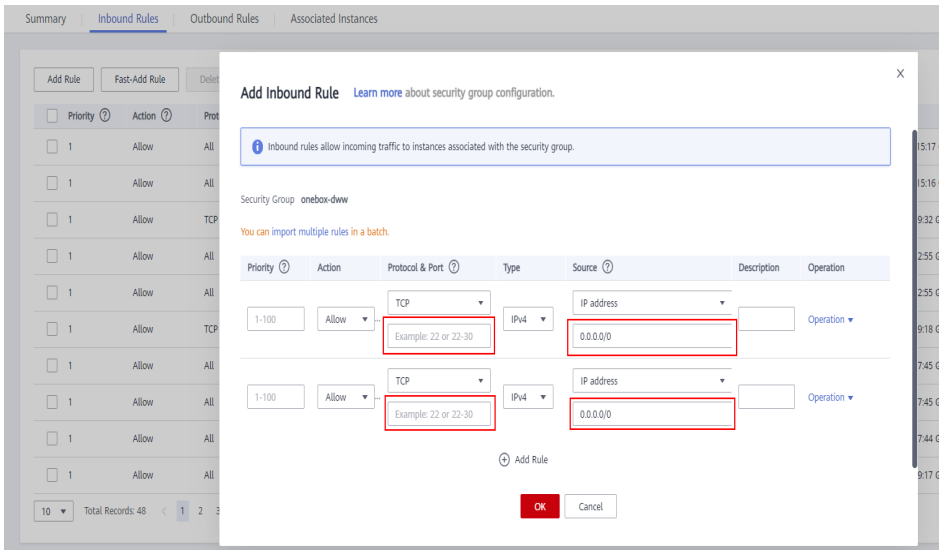


Step 2 Configure security group rules for the cluster.

1. On the **Dashboard** page, choose **Add Security Group Rule** > **Manage Security Group Rule**.



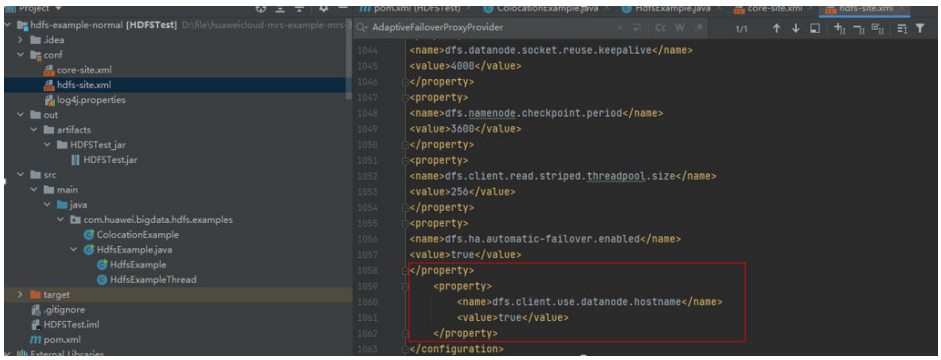
2. On the **Inbound Rules** tab page, click **Add Rule**. In the **Add Inbound Rule** dialog box, configure the Windows IP addresses and ports 8020 and 9866.



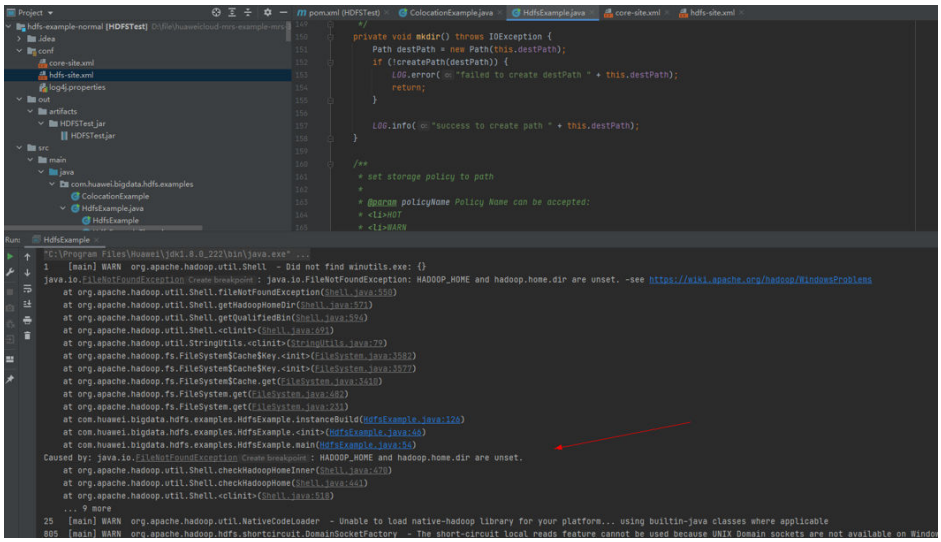
Step 3 On Manager, choose **Cluster** > **Services** > **HDFS** > **More** > **Download Client**, copy the **core-site.xml** and **hdfs-site.xml** files on the client to the **conf** directory of the sample project, and add the following content to the **hdfs-site.xml** file.

```
<property>
 <name>dfs.client.use.datanode.hostname</name>
 <value>true</value>
 </property>
```

(Change the DataNode communication mode to hostname.)



When you run the sample project, an error message indicating that hadoop_home does not exist may be displayed. You can ignore the error.



----End

6.7 How Do I Access Hive of the Cluster in Security Mode on Windows Using EIPs?

Scenario

This section describes how to bind Elastic IP addresses (EIPs) to a cluster and configure Hive files so that sample files can be compiled locally.

This section uses hive-jdbc-example as an example.

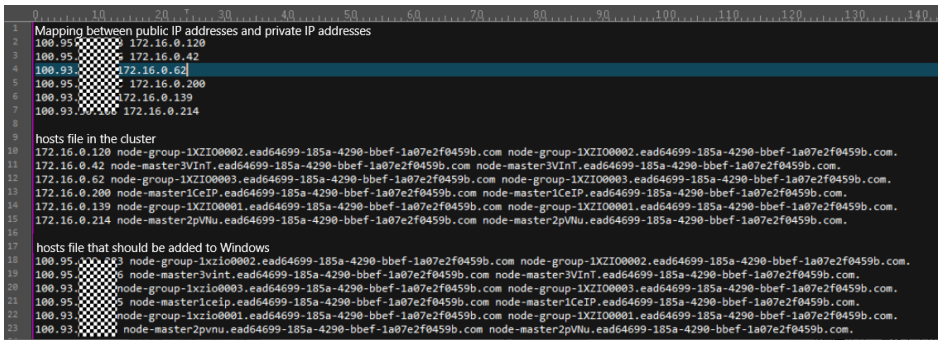
Procedure

Step 1 Apply for an EIP for each node in the cluster and add public IP addresses and corresponding host domain names of all nodes to the Windows local **hosts** file. (If a host name contains uppercase letters, change them to lowercase letters.)

1. On the VPC console, apply for EIPs (the number of EIPs you buy should be equal to the number of nodes in the cluster), click the name of each node in the MRS cluster, and bind an EIP to each node on the **EIPs** page.

   For details, see **Virtual Private Cloud** > **User Guide** > **EIP** > **Assigning an EIP and Binding It to an ECS**.

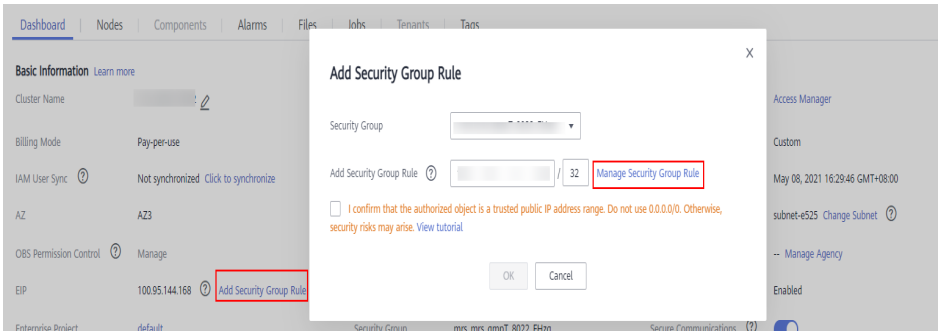
2. Record the mapping between the public IP addresses and private IP addresses. Change the private IP addresses in the **hosts** file to the corresponding public IP addresses.



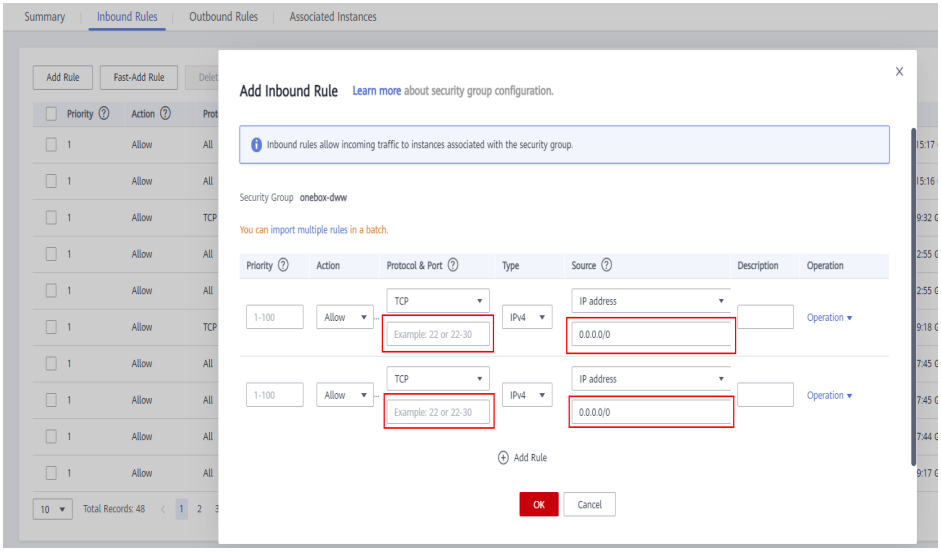
Step 2 Change the IP addresses in the **krb5.conf** file to the corresponding host names.

Step 3 Configure security group rules for the cluster.

1. On the **Dashboard** page, choose **Add Security Group Rule** > **Manage Security Group Rule**.



2. On the **Inbound Rules** tab page, click **Add Rule**. In the **Add Inbound Rule** dialog box, configure Windows IP addresses and ports 21730TCP, 21731TCP/UDP, and 21732TCP/UDP.

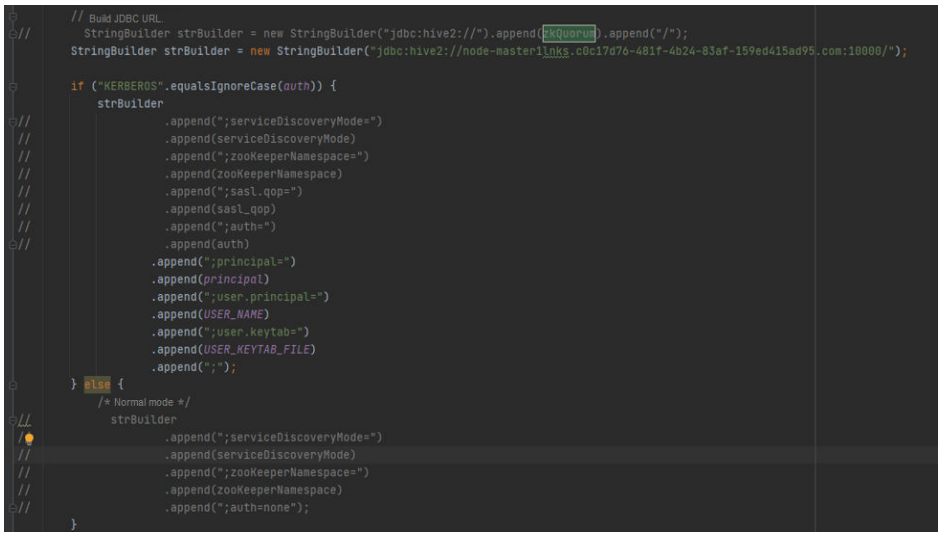


Step 4 On Manager, choose **Cluster** > **Services** > **Hive** > **More** > **Download Client**, and copy the **core-site.xml** and **hiveclient.properties** files on the client to the **resources** directory of the sample project.

Step 5 In the sample code, change the ZooKeeper IP addresses in the JDBC URL to the HiveServer2 host name for connection. Change the URL to **jdbc:hive2://** *HiveServer host name*:**10000/**.

NOTE

- The IP addresses in the **/hiveserver2** directory of ZooKeeper are private IP addresses and cannot be used to connect to Hive from Windows. Therefore, you need to replace ZooKeeper IP addresses with the HiveServer2 host name.
- To obtain the HiveServer2 host name, choose **Cluster** > **Services** > **Hive** > **Instance** on Manager and view **Host Name** of **HiveServer** on the **Instance** page.



Step 6 Before running the sample code, change **PRNCIPAL_NAME** in the sample code to the username for security authentication.

----End

6.8 How Do I Access Hive of the Cluster in Normal Mode on Windows Using EIPs?

Scenario

This section describes how to bind Elastic IP addresses (EIPs) to a cluster and configure Hive files so that sample files can be compiled locally.

This section uses hive-jdbc-example as an example.

Procedure

Step 1 Apply for an EIP for each node in the cluster and add public IP addresses and corresponding host domain names of all nodes to the Windows local **hosts** file. (If a host name contains uppercase letters, change them to lowercase letters.)

1. On the VPC console, apply for EIPs (the number of EIPs you buy should be equal to the number of nodes in the cluster), click the name of each node in the MRS cluster, and bind an EIP to each node on the **EIPs** page.

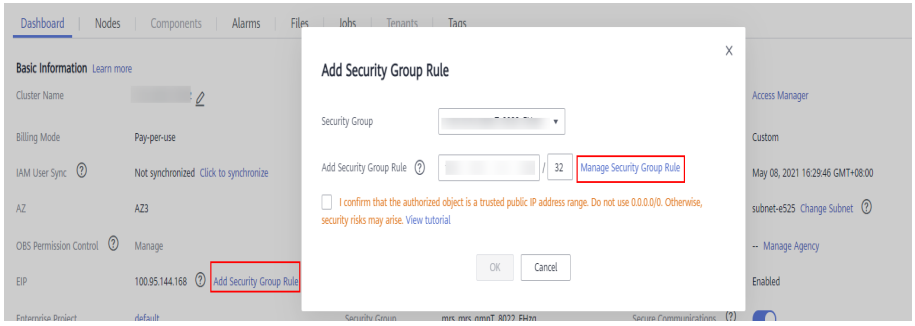
   For details, see **Virtual Private Cloud** > **User Guide** > **EIP** > **Assigning an EIP and Binding It to an ECS**.

2. Record the mapping between the public IP addresses and private IP addresses. Change the private IP addresses in the **hosts** file to the corresponding public IP addresses.

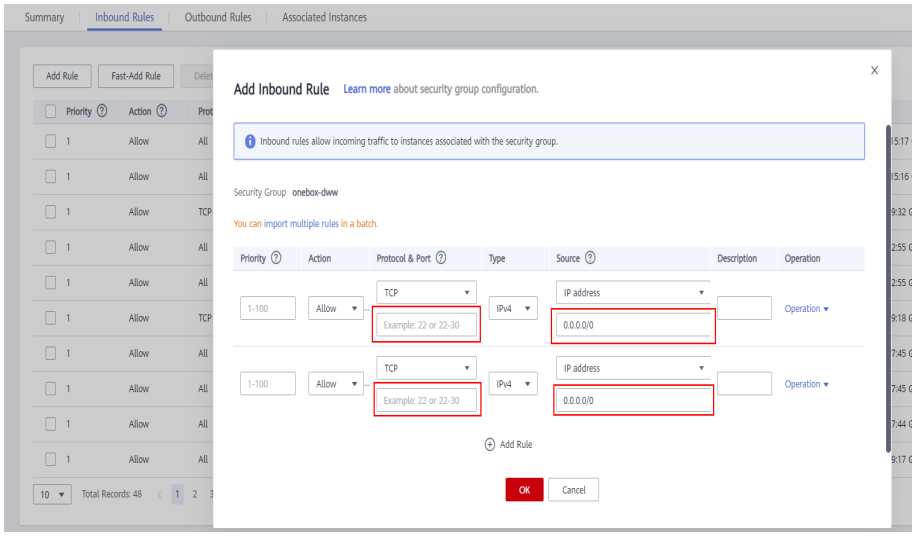
```
Mapping between public IP addresses and private IP addresses
100.95.[redacted] 172.16.0.120
100.95.[redacted] 172.16.0.42
100.93.[redacted] 172.16.0.62
100.95.[redacted] 172.16.0.200
100.93.[redacted] 172.16.0.139
100.93.[redacted] 172.16.0.214

hosts file in the cluster
172.16.0.120 node-group-1XZIO0002.ead64699-185a-4290-bbef-1a07e2f0459b.com node-group-1XZIO0002.ead64699-185a-4290-bbef-1a07e2f0459b.com.
172.16.0.42 node-master3VInT.ead64699-185a-4290-bbef-1a07e2f0459b.com node-master3VInT.ead64699-185a-4290-bbef-1a07e2f0459b.com.
172.16.0.62 node-group-1XZIO0003.ead64699-185a-4290-bbef-1a07e2f0459b.com node-group-1XZIO0003.ead64699-185a-4290-bbef-1a07e2f0459b.com.
172.16.0.200 node-master1CeIP.ead64699-185a-4290-bbef-1a07e2f0459b.com node-master1CeIP.ead64699-185a-4290-bbef-1a07e2f0459b.com.
172.16.0.139 node-group-1XZIO0001.ead64699-185a-4290-bbef-1a07e2f0459b.com node-group-1XZIO0001.ead64699-185a-4290-bbef-1a07e2f0459b.com.
172.16.0.214 node-master2pVNu.ead64699-185a-4290-bbef-1a07e2f0459b.com node-master2pVNu.ead64699-185a-4290-bbef-1a07e2f0459b.com.

hosts file that should be added to Windows
100.95.[redacted]3 node-group-1xzio0002.ead64699-185a-4290-bbef-1a07e2f0459b.com node-group-1XZIO0002.ead64699-185a-4290-bbef-1a07e2f0459b.com.
100.95.[redacted]6 node-master3vint.ead64699-185a-4290-bbef-1a07e2f0459b.com node-master3VInT.ead64699-185a-4290-bbef-1a07e2f0459b.com.
100.93.[redacted]node-group-1xzio0003.ead64699-185a-4290-bbef-1a07e2f0459b.com node-group-1XZIO0003.ead64699-185a-4290-bbef-1a07e2f0459b.com.
100.95.[redacted]5 node-master1ceip.ead64699-185a-4290-bbef-1a07e2f0459b.com node-master1CeIP.ead64699-185a-4290-bbef-1a07e2f0459b.com.
100.93.[redacted]node-group-1xzio0001.ead64699-185a-4290-bbef-1a07e2f0459b.com node-group-1XZIO0001.ead64699-185a-4290-bbef-1a07e2f0459b.com.
100.93.[redacted] node-master2pvnu.ead64699-185a-4290-bbef-1a07e2f0459b.com node-master2pVNu.ead64699-185a-4290-bbef-1a07e2f0459b.com.
```

Step 2 Configure security group rules for the cluster.

1. On the **Dashboard** page, choose **Add Security Group Rule** > **Manage Security Group Rule**.



2. On the **Inbound Rules** tab page, click **Add Rule**. In the **Add Inbound Rule** dialog box, configure the Windows IP address and port 10000.

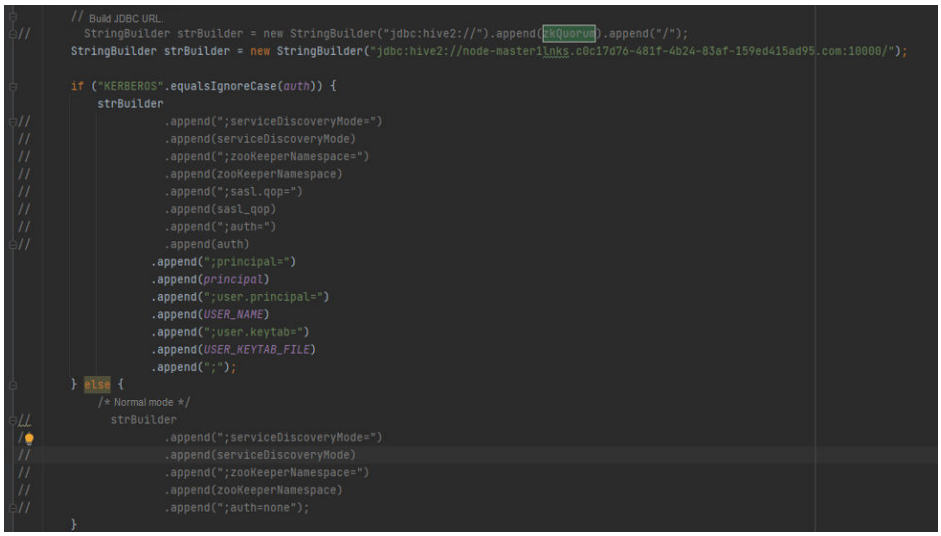


Step 3 On Manager, choose **Cluster** > **Services** > **Hive** > **More** > **Download Client**, and copy the **core-site.xml** and **hiveclient.properties** files on the client to the **resources** directory of the sample project.

Step 4 In the sample code, change the ZooKeeper IP addresses in the JDBC URL to the HiveServer2 host name for connection. Change the URL to **jdbc:hive2://** *HiveServer host name*:**10000/**.

NOTE

- The IP addresses in the **/hiveserver2** directory of ZooKeeper are private IP addresses and cannot be used to connect to Hive from Windows. Therefore, you need to replace ZooKeeper IP addresses with the HiveServer2 host name.
- To obtain the HiveServer2 host name, choose **Cluster** > **Services** > **Hive** > **Instance** on Manager and view **Host Name** of **HiveServer** on the **Instance** page.



----End

6.9 How Do I Access Kafka of the Cluster in Security Mode on Windows Using EIPs?

Scenario

This section describes how to bind Elastic IP addresses (EIPs) to a cluster and configure Kafka files so that sample files can be compiled locally.

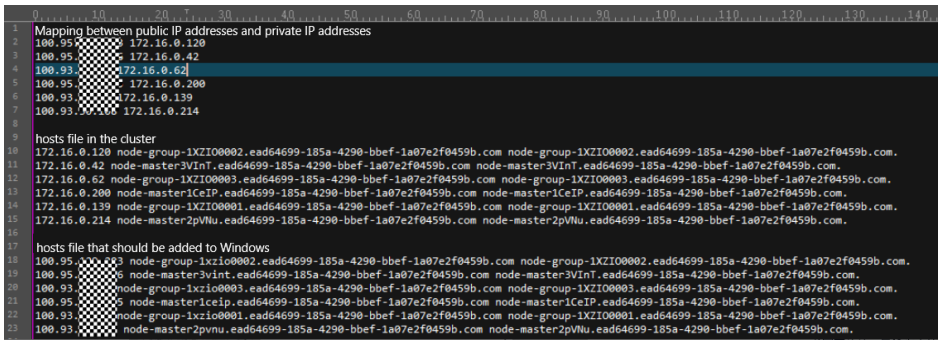
Procedure

Step 1 Apply for an EIP for each node in the cluster and add public IP addresses and corresponding host domain names of all nodes to the Windows local **hosts** file. (If a host name contains uppercase letters, change them to lowercase letters.)

1. On the VPC console, apply for EIPs (the number of EIPs you buy should be equal to the number of nodes in the cluster), click the name of each node in the MRS cluster, and bind an EIP to each node on the **EIPs** page.

   For details, see **Virtual Private Cloud** > **User Guide** > **EIP** > **Assigning an EIP and Binding It to an ECS**.

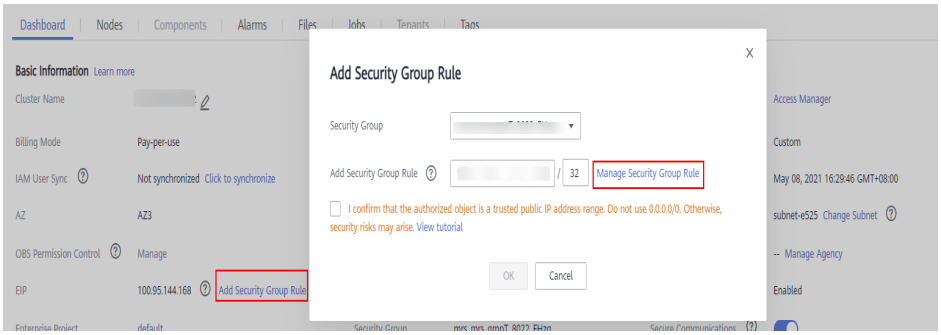
2. Record the mapping between the public IP addresses and private IP addresses. Change the private IP addresses in the **hosts** file to the corresponding public IP addresses.



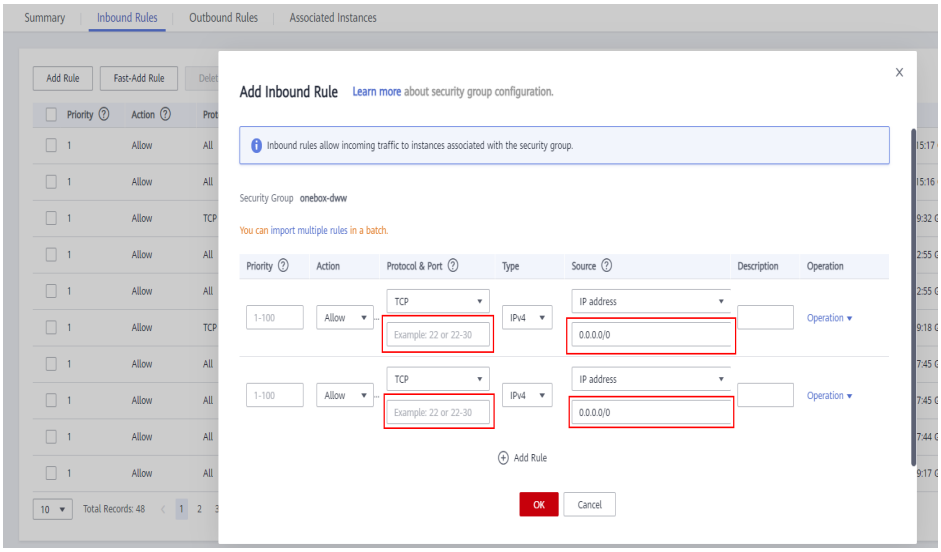
Step 2 Change the IP addresses in the **krb5.conf** file to the corresponding host names.

Step 3 Configure security group rules for the cluster.

1. On the **Dashboard** page, choose **Add Security Group Rule** > **Manage Security Group Rule**.



2. On the **Inbound Rules** tab page, click **Add Rule**. In the **Add Inbound Rule** dialog box, configure the Windows IP address and port 21007.



Step 4 On Manager, choose **Cluster** > **Services** > **Kafka** > **Configurations** > **All Configurations**, search for and add the key-value pair **advertised.listeners =SASL_PLAINTEXT://:21007,SASL_SSL://:21009,TRACE://:21013** in the **kafka.config.expandor** parameter, save the configuration, and restart the Kafka cluster.

Step 5 Before running the sample code, change the Kafka connection string in the sample code to **hostname1:21007**, **hostname2:21007**, **hostname3:21007**, change the

domain name in the code, and change the machine-machine account name and keytab file name applied by the user.

NOTE

You can log in to FusionInsight Manager, choose **System** > **Permission** > **Domain and Mutual Trust**, and check the value of **Local Domain**, which is the current system domain name.

----End

6.10 How Do I Access Kafka of the Cluster in Normal Mode on Windows Using EIPs?

Scenario

This section describes how to bind Elastic IP addresses (EIPs) to a cluster and configure Kafka files so that sample files can be compiled locally.

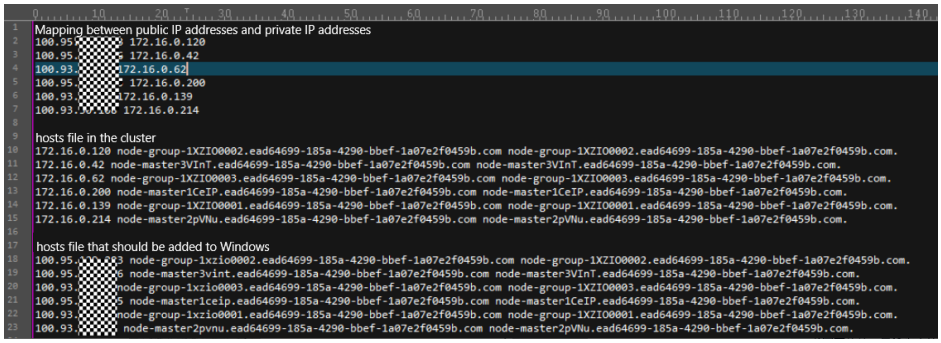
Procedure

Step 1 Apply for an EIP for each node in the cluster and add public IP addresses and corresponding host domain names of all nodes to the Windows local **hosts** file. (If a host name contains uppercase letters, change them to lowercase letters.)

1. On the VPC console, apply for EIPs (the number of EIPs you buy should be equal to the number of nodes in the cluster), click the name of each node in the MRS cluster, and bind an EIP to each node on the **EIPs** page.

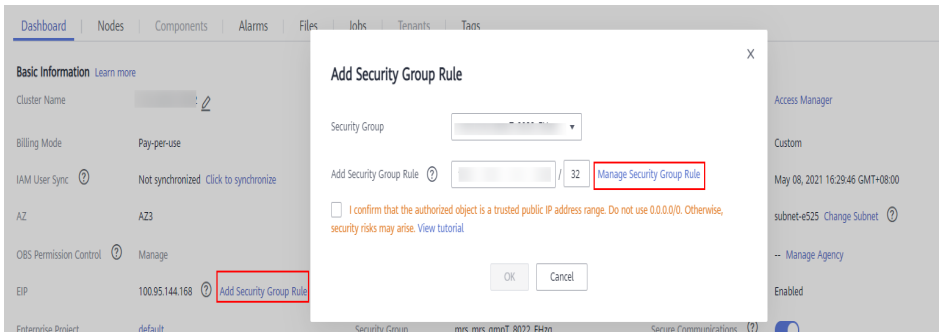
   For details, see **Virtual Private Cloud** > **User Guide** > **EIP** > **Assigning an EIP and Binding It to an ECS**.

2. Record the mapping between the public IP addresses and private IP addresses. Change the private IP addresses in the **hosts** file to the corresponding public IP addresses.

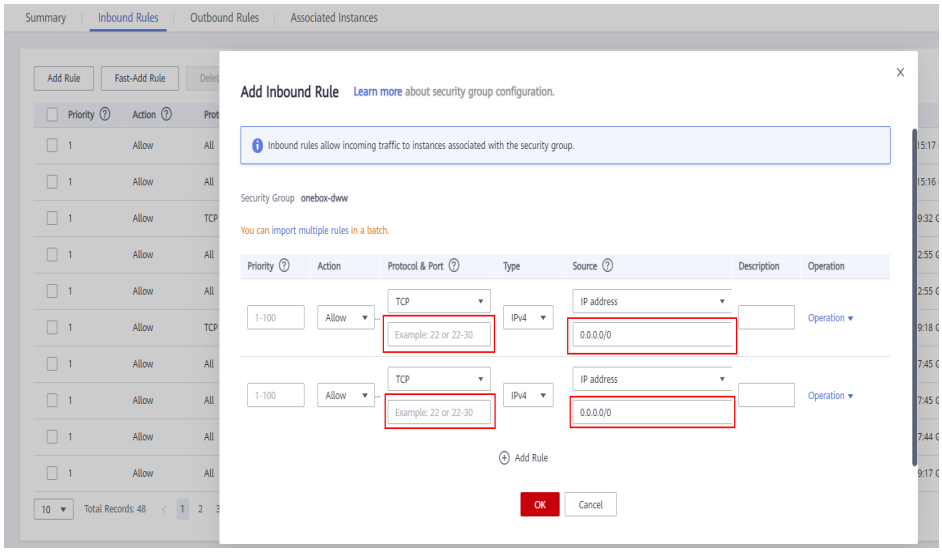


Step 2 Configure security group rules for the cluster.

1. On the **Dashboard** page, choose **Add Security Group Rule** > **Manage Security Group Rule**.



2. On the **Inbound Rules** tab page, click **Add Rule**. In the **Add Inbound Rule** dialog box, configure the Windows IP address and port 9092.



Step 3 On Manager, choose **Cluster** > **Services** > **Kafka** > **Configurations** > **All Configurations**, search for and add the key-value pair **advertised.listeners =PLAINTEXT://:9092,SSL://:9093,TRACE://:21013** in the **kafka.config.expandor** parameter, save the configuration, and restart the Kafka cluster.

Step 4 Before running the sample code, change the Kafka connection string in the sample code to *hostname1*:**9092,** *hostname2*:**9092,** *hostname3*:**9092**.

----End

6.11 How Do I Access Spark of the Cluster in Security Mode on Windows Using EIPs?

Scenario

This section describes how to bind Elastic IP addresses (EIPs) to a cluster and configure Spark files so that sample files can be compiled locally.

This section uses SparkScalaExample as an example.

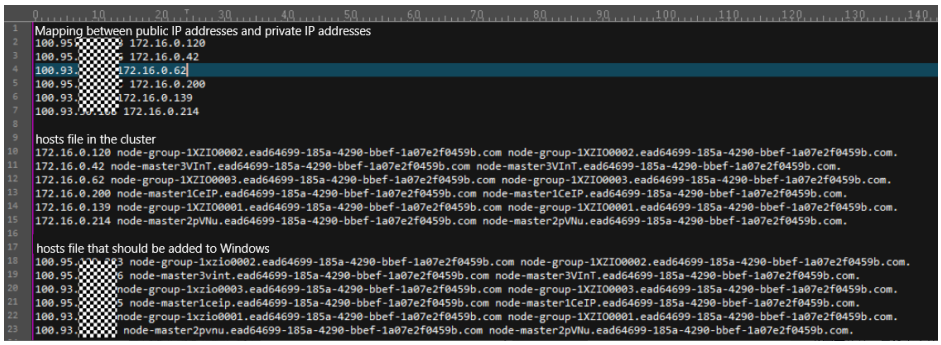
Procedure

Step 1 Apply for an EIP for each node in the cluster and add public IP addresses and corresponding host domain names of all nodes to the Windows local **hosts** file. (If a host name contains uppercase letters, change them to lowercase letters.)

1. On the VPC console, apply for EIPs (the number of EIPs you buy should be equal to the number of nodes in the cluster), click the name of each node in the MRS cluster, and bind an EIP to each node on the **EIPs** page.

   For details, see **Virtual Private Cloud** > **User Guide** > **EIP** > **Assigning an EIP and Binding It to an ECS**.

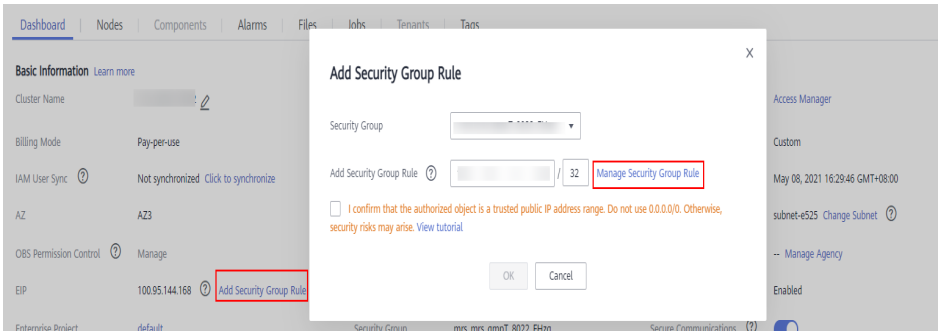
2. Record the mapping between the public IP addresses and private IP addresses. Change the private IP addresses in the **hosts** file to the corresponding public IP addresses.



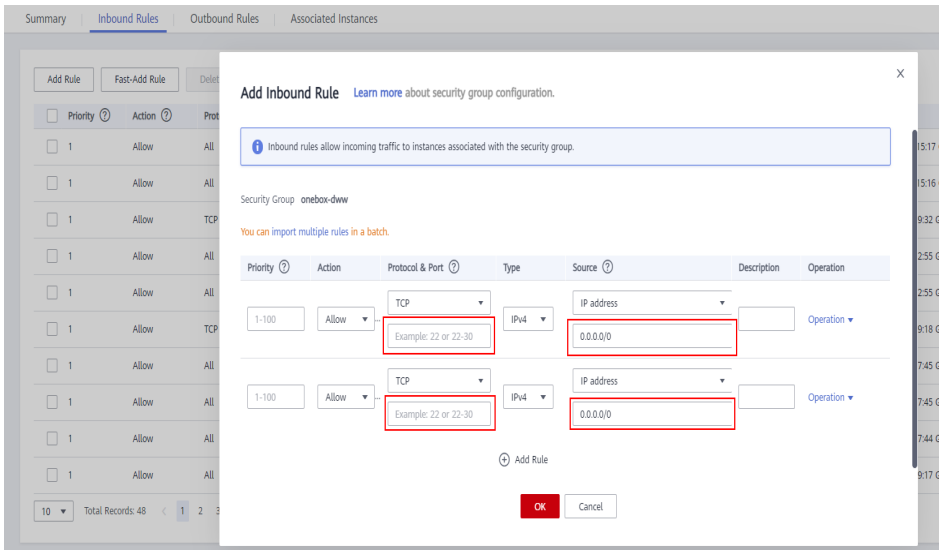
Step 2 Change the IP addresses in the **krb5.conf** file to the corresponding host names.

Step 3 Configure security group rules for the cluster.

1. On the **Dashboard** page, choose **Add Security Group Rule** > **Manage Security Group Rule**.



2. On the **Inbound Rules** tab page, click **Add Rule**. In the **Add Inbound Rule** dialog box, configure Windows IP addresses and ports 21730TCP, 21731TCP/UDP, and 21732TCP/UDP.



Step 4 On Manager, choose **Cluster** > **Services** > **HDFS** > **More** > **Download Client**, and copy the **core-site.xml** and **hdfs-site.xml** files on the client to the **conf** directory of the sample project.

Add the following content to the **hdfs-site.xml** file:

```
<property>
      <name>dfs.client.use.datanode.hostname</name>
      <value>true</value>
</property>
```

Add the following content to the **pom.xml** file:

```
<dependency>
    <groupId>com.huawei.mrs</groupId>
    <artifactId>hadoop-plugins</artifactId>
    <version>Component package version-302002</version>
</dependency>
```

Step 5 Before running the sample code, add **.master("local").config("spark.driver.host", "localhost")** to **SparkSession** to set the local running mode for Spart. Change **PRNCIPAL_NAME** in the sample code to the username for security authentication.

```
object FemaleInfoCollection {
  def main (args: Array[String]) {
//    if (args.length < 1) {
//      System.err.println("Usage: CollectFemaleInfo <file>")
//      System.exit(1)
//    }

    // Configure the Spark application name.
    val spark = SparkSession
      .builder()
      .appName( name = "CollectFemaleInfo")
      .master( master = "local")
      .config("spark.driver.host", "localhost")
      .getOrCreate()
    // Initializing Spark


    // Read data. This code indicates the data path that the input parameter args(0) specifies.
    val text = spark.sparkContext.textFile( path = "/tmp/sparktest/")
    // Filter the data information about the time that female netizens spend online.
    val data = text.filter(_.contains("female"))
```

----End

6.12 How Do I Access Spark of the Cluster in Normal Mode on Windows Using EIPs?

Scenario

This section describes how to bind Elastic IP addresses (EIPs) to a cluster and configure Spark files so that sample files can be compiled locally.

This section uses SparkScalaExample as an example.

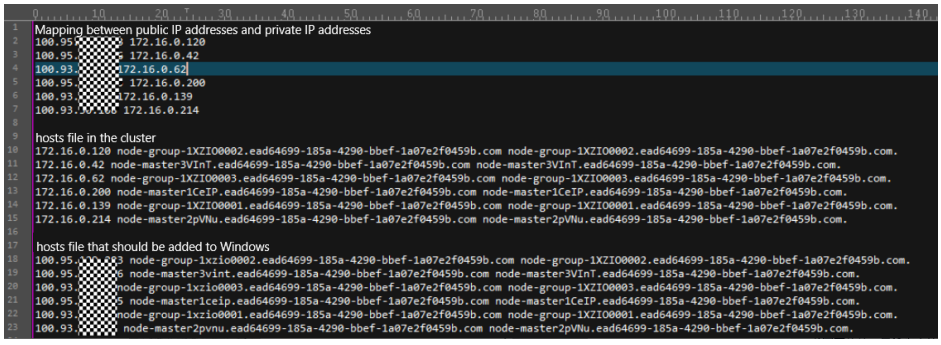
Procedure

Step 1 Apply for an EIP for each node in the cluster and add public IP addresses and corresponding host domain names of all nodes to the Windows local **hosts** file. (If a host name contains uppercase letters, change them to lowercase letters.)

1. On the VPC console, apply for EIPs (the number of EIPs you buy should be equal to the number of nodes in the cluster), click the name of each node in the MRS cluster, and bind an EIP to each node on the **EIPs** page.

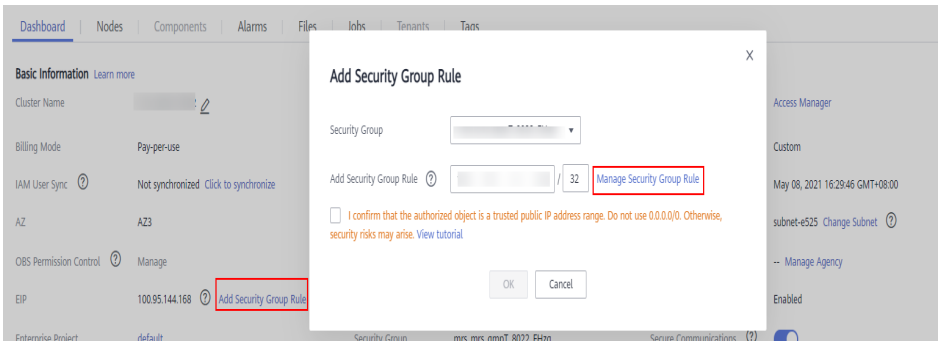
   For details, see **Virtual Private Cloud** > **User Guide** > **EIP** > **Assigning an EIP and Binding It to an ECS**.

2. Record the mapping between the public IP addresses and private IP addresses. Change the private IP addresses in the **hosts** file to the corresponding public IP addresses.

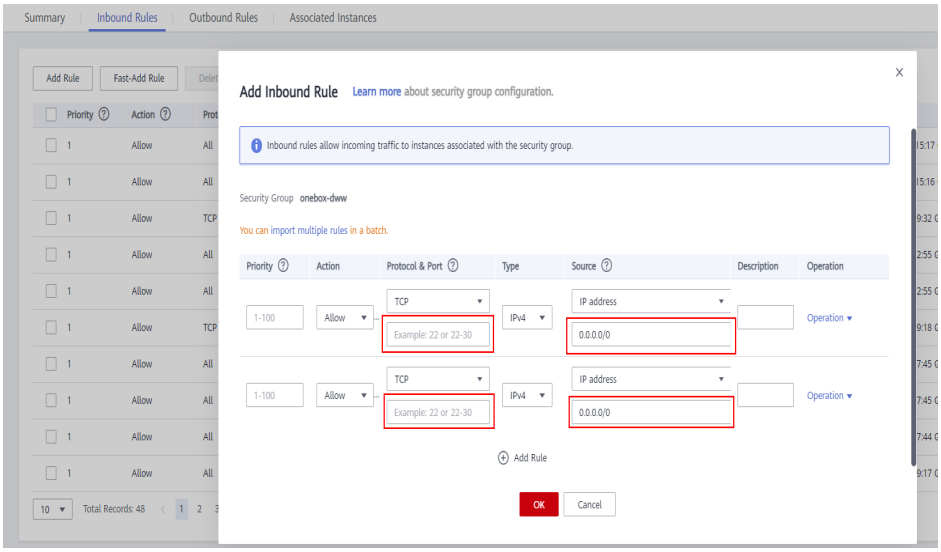


Step 2 Configure security group rules for the cluster.

1. On the **Dashboard** page, choose **Add Security Group Rule** > **Manage Security Group Rule**.



2. On the **Inbound Rules** tab page, click **Add Rule**. In the **Add Inbound Rule** dialog box, configure the Windows IP addresses and ports 8020 and 9866.



Step 3 On Manager, choose **Cluster** > **Services** > **HDFS** > **More** > **Download Client**, and copy the **core-site.xml** and **hdfs-site.xml** files on the client to the **conf** directory of the sample project.

Add the following content to the **hdfs-site.xml** file:

```
<property>
      <name>dfs.client.use.datanode.hostname</name>
      <value>true</value>
</property>
```

Add the following content to the **pom.xml** file:

```
<dependency>
    <groupId>com.huawei.mrs</groupId>
    <artifactId>hadoop-plugins</artifactId>
    <version>Component package version-302002</version>
</dependency>
```

Step 4 Before running the sample code, add **.master("local").config("spark.driver.host", "localhost")** to **SparkSession** to set the local running mode for Spart.

```
object FemaleInfoCollection {
  def main (args: Array[String]) {
//    if (args.length < 1) {
//      System.err.println("Usage: CollectFemaleInfo <file>")
//      System.exit(1)
//    }

    // Configure the Spark application name.
    val spark = SparkSession
      .builder()
      .appName( name = "CollectFemaleInfo")
      .master( master = "local")
      .config("spark.driver.host", "localhost")
      .getOrCreate()
    // Initializing Spark


    // Read data. This code indicates the data path that the input parameter args(0) specifies.
    val text = spark.sparkContext.textFile( path = "/tmp/sparktest/")
    // Filter the data information about the time that female netizens spend online.
    val data = text.filter(_.contains("female"))
```

----End

6.13 How Do I Access HBase of the Cluster in Security Mode on Windows Using EIPs?

Scenario

This section describes how to bind Elastic IP addresses (EIPs) to a cluster and configure HBase files so that sample files can be compiled locally.

This section uses hbase-example as an example.

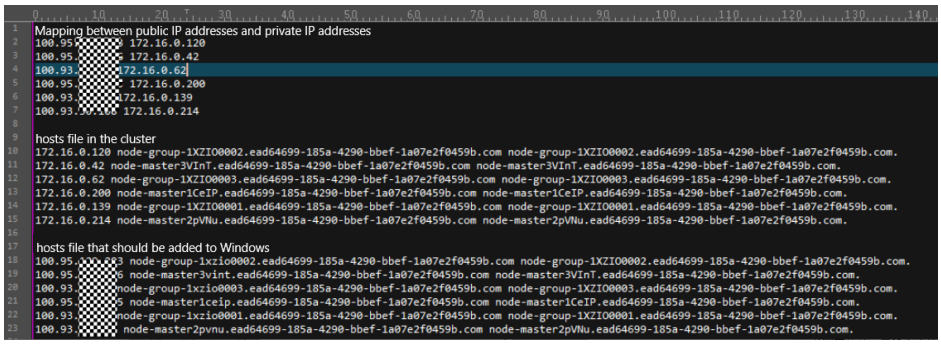
Procedure

Step 1 Apply for an EIP for each node in the cluster and add public IP addresses and corresponding host domain names of all nodes to the Windows local **hosts** file. (If a host name contains uppercase letters, change them to lowercase letters.)

1. On the VPC console, apply for EIPs (the number of EIPs you buy should be equal to the number of nodes in the cluster), click the name of each node in the MRS cluster, and bind an EIP to each node on the **EIPs** page.

   For details, see **Virtual Private Cloud** > **User Guide** > **EIP** > **Assigning an EIP and Binding It to an ECS**.

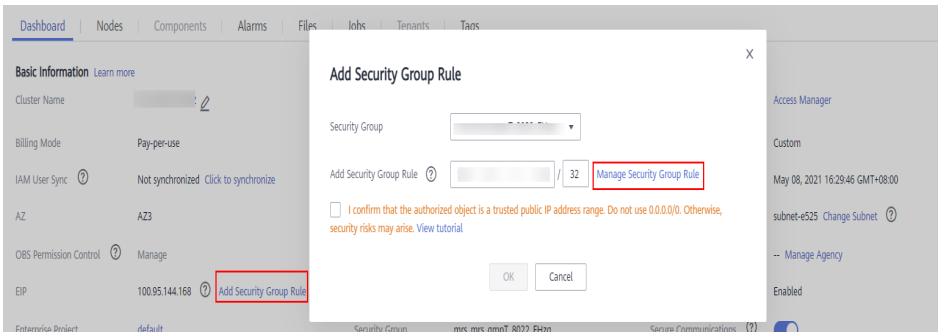
2. Record the mapping between the public IP addresses and private IP addresses. Change the private IP addresses in the **hosts** file to the corresponding public IP addresses.



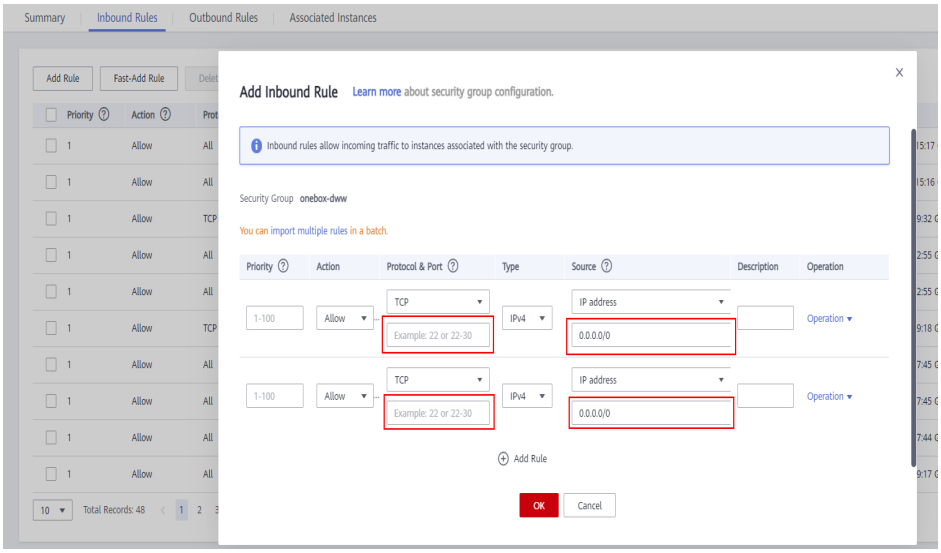
Step 2 Change the IP addresses in the **krb5.conf** file to the corresponding host names.

Step 3 Configure security group rules for the cluster.

1. On the **Dashboard** page, choose **Add Security Group Rule** > **Manage Security Group Rule**.



2. On the **Inbound Rules** page, click **Add Rule**. On the displayed **Add Inbound Rule** page, configure Windows IP addresses and ports **21730TCP**, **21731TCP/UDP**, and **21732TCP/UDP**.



Step 4 On Manager, choose **Cluster** > **Services** > **HBase** > **More** > **Download Client**, and copy the **core-site.xml**, **hdfs-site.xml**, and **hbase-site.xml** files on the client to the **resources** directory of the sample project.

Step 5 Before running the sample code, change the value of **userName** in the sample code to the username for security authentication.

----End

6.14 How Do I Access HBase of the Cluster in Normal Mode on Windows Using EIPs?

Scenario

This section describes how to bind Elastic IP addresses (EIPs) to a cluster and configure HBase files so that sample files can be compiled locally.

This section uses hbase-example as an example.

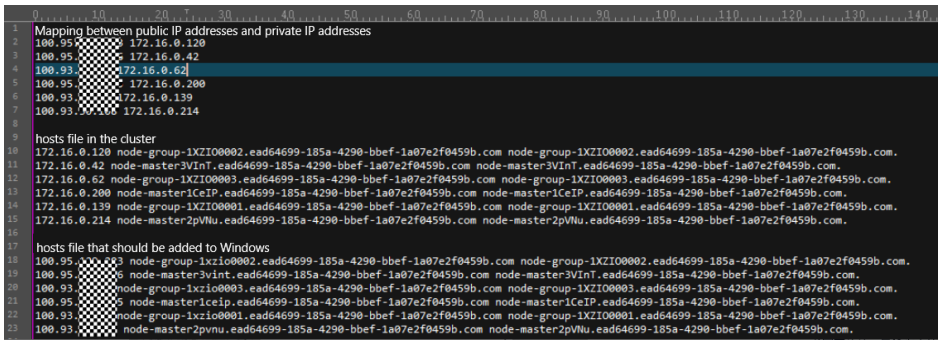
Procedure

Step 1 Apply for an EIP for each node in the cluster and add public IP addresses and corresponding host domain names of all nodes to the Windows local **hosts** file. (If a host name contains uppercase letters, change them to lowercase letters.)

1. On the VPC console, apply for EIPs (the number of EIPs you buy should be equal to the number of nodes in the cluster), click the name of each node in the MRS cluster, and bind an EIP to each node on the **EIPs** page.

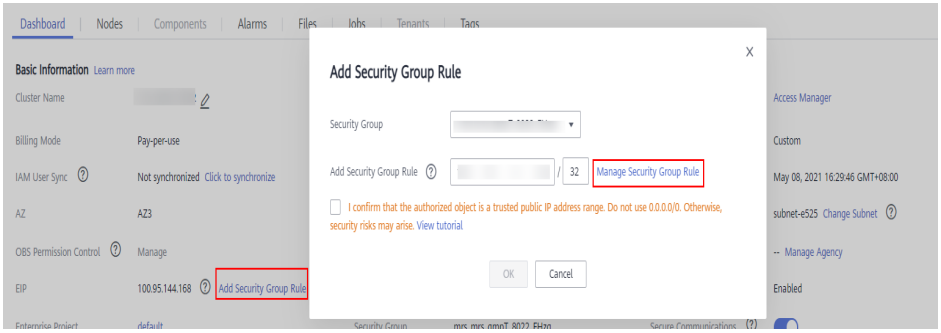
   For details, see **Virtual Private Cloud** > **User Guide** > **EIP** > **Assigning an EIP and Binding It to an ECS**.

2. Record the mapping between the public IP addresses and private IP addresses. Change the private IP addresses in the **hosts** file to the corresponding public IP addresses.

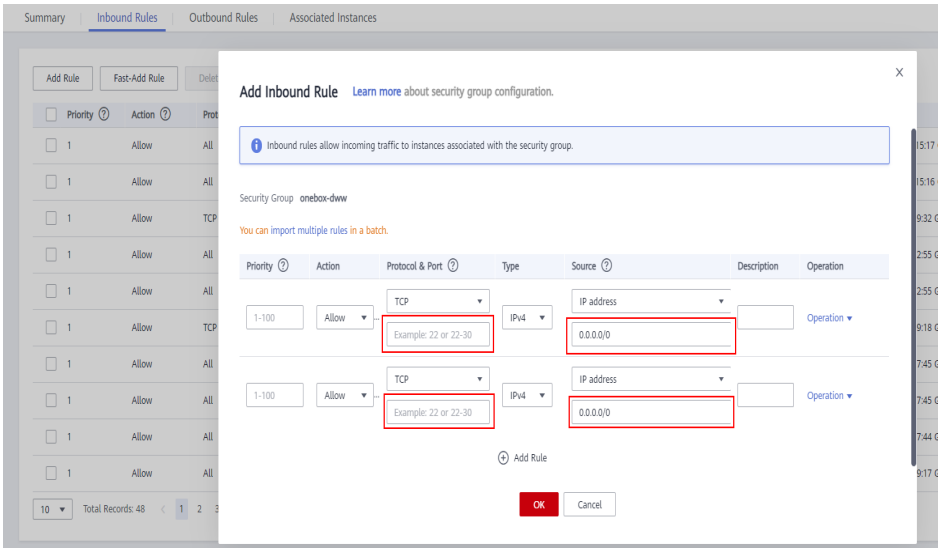


Step 2 Configure security group rules for the cluster.

1. On the **Dashboard** page, choose **Add Security Group Rule** > **Manage Security Group Rule**.



2. On the **Inbound Rules** page, click **Add Rule**. On the displayed **Add Inbound Rule** page, configure Windows IP addresses and ports **21730TCP**, **21731TCP/UDP**, and **21732TCP/UDP**.



Step 3 On Manager, choose **Cluster** > **Services** > **HBase** > **More** > **Download Client**, and copy the **core-site.xml**, **hdfs-site.xml**, and **hbase-site.xml** files on the client to the **resources** directory of the sample project.

----End

6.15 How Do I Switch the Mode of Accessing MRS Manager?

Question

How do I use an EIP to access MRS Manager after I access MRS Manager using a Direct Connect connection?

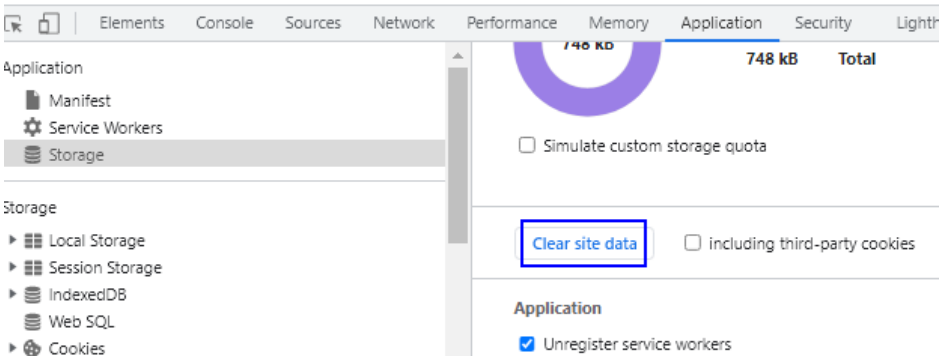
Answer

- For MRS 3.*x* or later:

  On the **Dashboard** tab page of the cluster, click next to **Access Manager** to switch the access mode.

- For MRS 2.*x* or earlier:

  a. On the cluster page, press **F12**, choose **Application** > **Storage**, and click **Clear site data**.

  b. Refresh the page and log in to the MRS cluster page again. On the **Dashboard** tab page, click **Access Manager** to switch the access mode.

7 Alarm Monitoring

7.1 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.

7.2 Where Can I View the Running Resource Queues When the Alarm "ALM-18022 Insufficient Yarn Queue Resources" Is Reported?

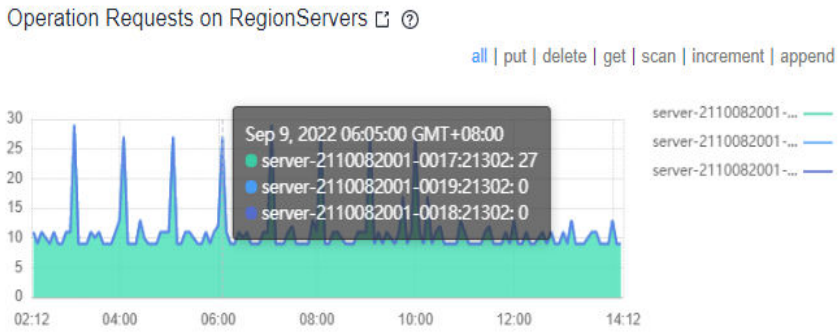
Log in to FusionInsight Manager and choose **Cluster** > **Services** > **Yarn**. In the navigation pane on the left, choose **ResourceManager(Active)** and log in to the native Yarn page.

For details, see the online help.

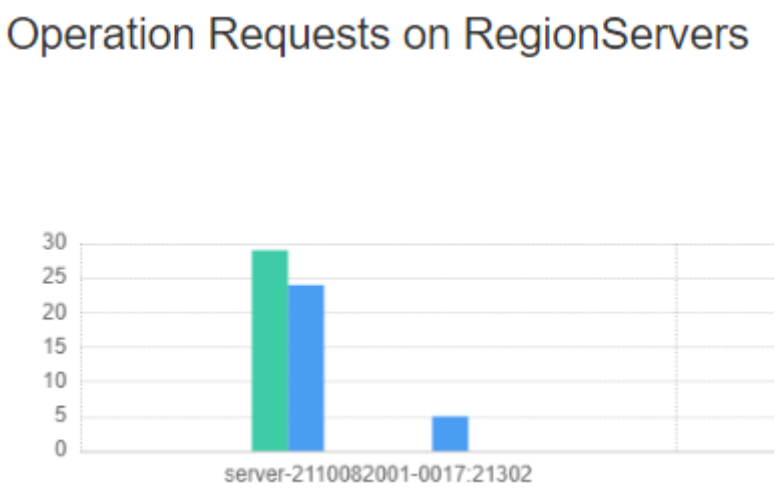
7.3 How Do I Understand the Multi-Level Chart Statistics in the HBase Operation Requests Metric?

The following uses the **Operation Requests on RegionServers** monitoring item as an example:

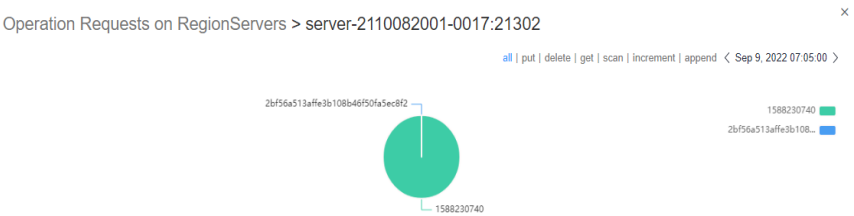
1. Log in to FusionInsight Manager and choose **Cluster** > **Services** > **HBase** > **Resource**. On the displayed page, you can view the **Operation Requests on RegionServers** chart. If you click **all**, the top 10 RegionServers ranked by the total number of operation requests in the current cluster are displayed, the statistics interval is 5 minutes.



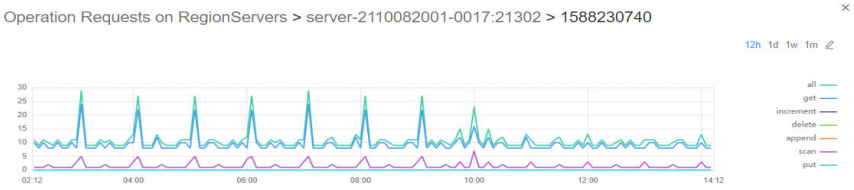
2. Click a point in the chart. A level-2 chart is displayed, showing the number of operation requests of all RegionServers in the past 5 minutes.



3. Click an operation statistics bar chart. A level-3 chart is displayed, showing the distribution of operations in each region within the period.



4. Click a region name. The distribution chart of operations performed every 5 minutes in the last 12 hours is displayed. You can view the number of operations performed in the period.



8 Performance Tuning

8.1 Does an MRS Cluster Support System Reinstallation?

An MRS cluster does not support system reinstallation.

8.2 Can I Change the OS of an MRS Cluster?

The OS of an MRS cluster cannot be changed.

8.3 How Do I Improve the Resource Utilization of Core Nodes in a Cluster?

1. Go to the Yarn service configuration page.
   - For MRS 1.8.10 or earlier clusters:

     Log in to MRS Manager by referring to **Accessing MRS Manager**. Choose **Services** > **Yarn** > **Service Configuration**, and select **All** from the **Basic** drop-down list.
   - For MRS 1.8.10 or later and MRS 2.*x*, click the cluster name. On the cluster details page that is displayed, click the **Components** tab and choose **Yarn**. Click the **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 cluster version 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.

8.4 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

9 Job Development

9.1 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.

9.2 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.

9.3 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.

9.4 How Do I Do If Job Parameters Separated By Spaces Cannot Be Identified?

Use spaces to separate parameters. To prevent parameters from being saved as plaintexts, add an at sign (@) before parameters, for example, **@password=admin_123**.

9.5 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.

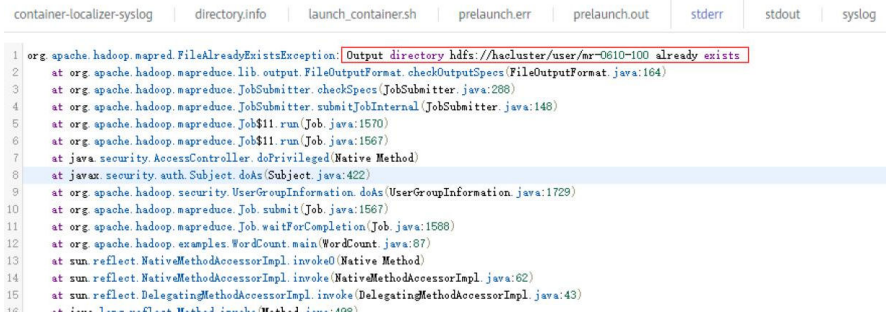
In Yarn-cluster mode, Driver runs in ApplicationMaster, which requests resources from Yarn and monitors the running status of a job. After a user submits a job, the client can be stopped and the job continues running on Yarn. Therefore, the Yarn-cluster mode is not suitable for running interactive jobs.

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.

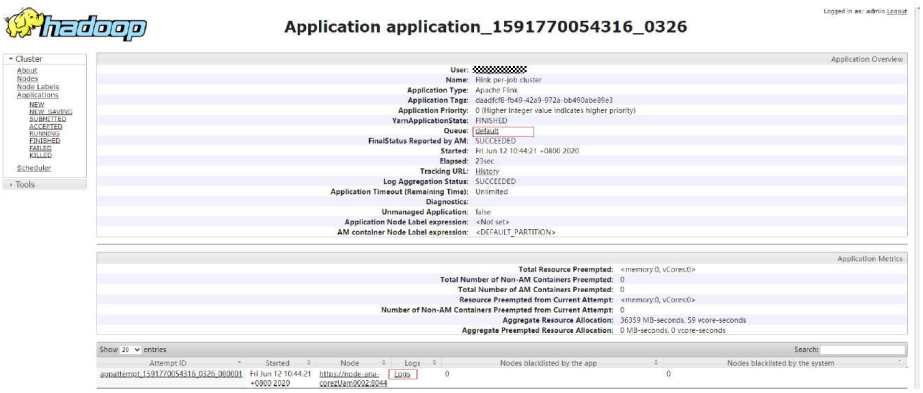
9.6 How Do I View MRS Job Logs?

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:



- You can view realJob logs on the ResourceManager web UI provided by the Yarn service on MRS Manager.



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

9.7 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.

9.8 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.

9.9 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.

9.10 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.

9.11 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
  ```

9.12 How Do I Do If Error Message "java.io.IOException: Connection reset by peer" Is Displayed During the Execution of a Spark Job?

Symptom

The Spark job keeps running and error message "java.io.IOException: Connection reset by peer" is displayed.

Solution

Add the **executor.memory Overhead** parameter to the parameters for submitting a job.

9.13 How Do I Do If Error Message "requestId=49718838510717372 50" Is Displayed When a Spark Job Accesses OBS?

Symptom

Error message "requestId=4971883851071737250" is displayed when a Spark job accesses OBS.

Solution

Log in to the node where the Spark client is located, go to the **conf** directory, and change the value of the **fs.obs.metrics.switch** parameter in the **core-site.xml** configuration file to **false**.

9.14 How Do I Do If the Spark Job Error "UnknownScannerExeception" Is Reported?

Symptom

Spark jobs run slowly. Warning information is printed in run logs, and the error cause is **UnknownScannerExeception**.

Solution

Before running a Spark job, adjust the value of **hbase.client.scanner.timeout.period** (for example, from 60 seconds to 120 seconds).

Log in to FusionInsight Manager and choose **Cluster** > **Services** > **HBase**. Click **Configurations** then **All Configurations**, search for **hbase.client.scanner.timeout.period**, and change its value to **120000** (unit: ms).

9.15 Why DataArtsStudio Occasionally Fail to Schedule Spark Jobs and the Rescheduling also Fails?

Symptom

DataArtsStudio occasionally fails to schedule Spark jobs and the rescheduling also fails. The following error information is displayed:

```
Caused by: org.apache.spark.SparkException: Application application_1619511926396_2586346 finished with failed status
```


Solution

Log in to the node where the Spark client is located as user **root** and increase the value of the **spark.driver.memory** parameter in the **spark-defaults.conf** file.

9.16 How Do I Do If a Flink Job Fails to Execute and the Error Message "java.lang.NoSuchFieldError: SECURITY_SSL_ENCRYPT_ENABLED" Is Displayed?

Symptom

A Flink job fails to be executed and the following error message is displayed:

```
Caused by: java.lang.NoSuchFieldError: SECURITY_SSL_ENCRYPT_ENABLED
```

Solution

The third-party dependency package in the customer code conflicts with the cluster package. As a result, the job fails to be submitted to the MRS cluster. You need to modify the dependency package, set the scope of the open source Hadoop package and Flink package in the POM file to **provide**, and pack and execute the job again.

9.17 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.

9.18 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.

9.19 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.

9.20 How Do I Do If the Error Message "slot request timeout" Is Displayed When I Submit a Flink Job?

Symptom

When a Flink job is submitted, JobManager is started successfully. However, TaskManager remains in the starting state until timeout. The following error information is displayed:

```
org.apache.flink.runtime.jobmanager.scheduler.NoResourceAvailableException: Could not allocate the required slot within slot request timeout. Please make sure that the cluster has enough resources
```

Possible Causes

1. The resources in the YARN queue are insufficient. As a result, TaskManager fails to start.
2. Your JAR files conflict with those in the environment. You can execute the WordCount program to determine whether the issue occurs.
3. If the cluster is in security mode, the SSL certificate of Flink may be incorrectly configured or has expired.

Solution

1. Add resources to the YARN queue.
2. Exclude the Flink and Hadoop dependencies in your JAR files so that Flink and Hadoop can depend only on the JAR files in the environment.
3. Reconfigure the SSL certificate of Flink..

9.21 Data Import and Export of DistCP Jobs

- Does a DistCP job compare data consistency during data import and export?

  No. DistCP jobs only copy data but do not modify it.

- When data is exported from a DistCP job, if some files already exist in OBS, how will the job process the files?

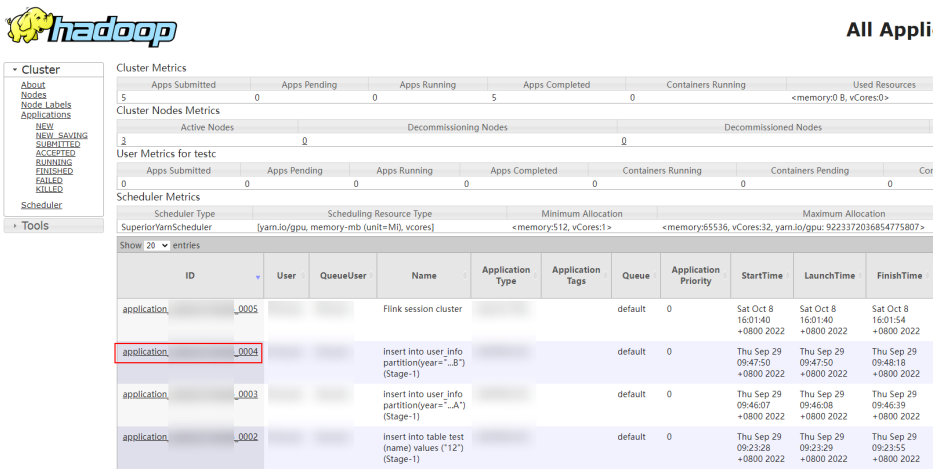
  DistCP jobs will overwrite the files in OBS.

9.22 How Do I View SQL Statements for Hive Jobs on the YARN Web UI?

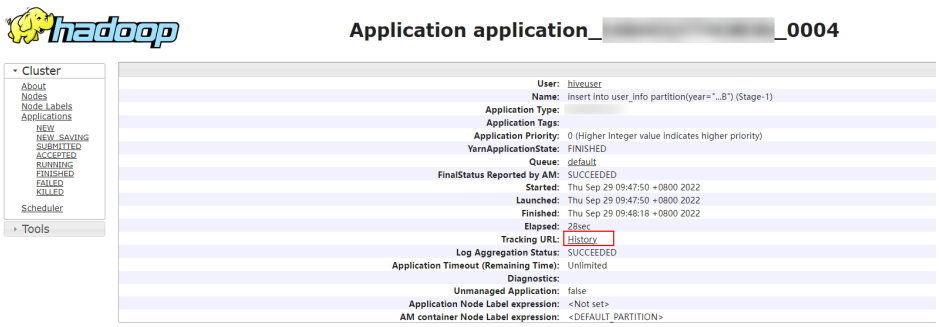
Step 1 Log in to FusionInsight Manager as a service user.

Step 2 Choose **Cluster** > **Services** > **Yarn**. Click **ResourceManager(xxx,Active)** next to **ResourceManager WebUI**.

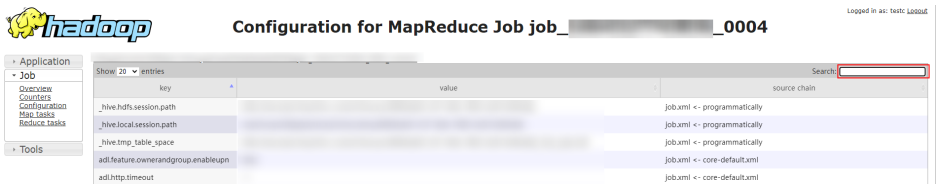
Step 3 On the YARN web UI, click the ID of the job to be viewed.



Step 4 Click **ApplicationMaster** or **History** next to **Tracking URL**.



Step 5 Choose **Configuration** in the left navigation pane and search for the **hive.query.string** parameter in the upper right corner to query the corresponding HiveSQL.



----End

10 Cluster Upgrade/Patching

10.1 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.

10.2 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.

11 Peripheral Ecosystem Interconnection

11.1 Can MRS Be Used to Perform Read and Write Operations on DLI Tables?

If you have stored data on OBS, you can use Spark in MRS to read Data Lake Insight (DLI) tables, flexibly process the table data, and save the result to another DLI table. If you have not stored data on OBS, you cannot use MRS to read or write DLI tables.

11.2 Does OBS Support the ListObjectsV2 Protocol?

OBS does not support the ListObjectsV2 protocol.

11.3 Can MRS Data Be Stored in a Parallel File System Provided by OBS?

MRS supports the storage on a parallel file system provided by OBS.

11.4 Can the Crawler Service Be Deployed in MRS?

The Crawler Service cannot be deployed in MRS.

11.5 Do DWS and MRS Support Secure Deletion (Preventing Retrieval After Deletion)?

Both MRS and Data Warehouse Service (DWS) support secure deletion. MRS supports restoration only from the backup data. DWS supports restoration only from snapshots.

11.6 Why Is the Kerberos-Authenticated MRS Cluster Not Found When a Connection Is Set Up from DLF?

Data Lake Factory (DLF) cannot interconnect with an MRS cluster that has Kerberos authentication enabled. If you want to use MRS clusters with Kerberos authentication enabled, use the DataArts Studio platform.

11.7 How Do I Use PySpark on an ECS to Connect to an MRS Spark Cluster with Kerberos Authentication Enabled, on the Intranet?

Change the value of **spark.yarn.security.credentials.hbase.enabled** in the **spark-defaults.conf** file of Spark to **true** and use **spark-submit --master yarn --keytab keytabfile --principal principal** command to specify the Kerberos authentication file.

11.8 Why Mapped Fields Do not Exist in the Database After HBase Synchronizes Data to CSS?

After data is synchronized from HBase of MRS to Cloud Search Service (CSS), the database will not have mapped fields. Only tables have mapped fields.

11.9 Can Flume Read Data from OBS?

By default, Flume cannot read data from OBS. If you have service requirements, you need to customize sources to implement this function.

11.10 Can MRS Connect to an External KDC?

MRS supports only the built-in key distribution center (KDC).

11.11 How Do I Solve the Jetty Version Compatibility Issue in Open-Source Kylin 3.x and MRS 1.9.3 Interconnection?

For security purposes, MRS has upgraded some open-source third-party components that have serious security vulnerabilities. The upgrade causes the Jetty version compatibility issue when the open-source Kylin interconnects with MRS 1.9.3.

Perform the following operations to solve the issue:

1. Install the MRS client on an ECS node. . This operation uses the MRS client installation path **/srv/client/** as an example.

2. After the installation is complete, run the following commands to import the MRS client environment variable **bigdata_env** and the environment variables **HIVE_CONF** and **HCAT_HOME** required by Kylin.

   source /srv/client/bigdata_env

   export HIVE_CONF=/srv/client/Hive/config/

   export HCAT_HOME=/srv/client/Hive/HCatalog

3. Install Kylin on the node where the MRS client is installed and specify **KYLIN_HOME**. For details, see the **Kylin official website**. For MRS 1.9.3, select Kylin for HBase 1.*x* for interconnection.

   export KYLIN_HOME=/srv/client/apache-kylin-3.0.2-bin-hbase1x

4. Remove Jetty .jar packages from the **/srv/client/Hive/Beeline/lib/** directory in the Hive client directory to prevent version conflicts.

   Jetty .jar packages:

   ```
   javax-websocket-server-impl-9.4.26.v20200117.jar
   websocket-server-9.4.26.v20200117.jar
   jetty-all-9.4.26.v20200117-uber.jar
   jetty-runner-9.4.26.v20200117.jar
   apache-jsp-9.4.26.v20200117.jar
   ```

5. Start the Kylin service and check Kylin logs. In normal cases, the logs do not contain compatibility errors, such as **java.lang.NoSuchMethodException** and **java.lang.ClassNotFoundException**.

   $KYLIN_HOME/bin/kylin.sh start

6. Access the native Kylin page at **http://<hostname>:7070/kylin** and run the sample Cube script **${KYLIN_HOME}/bin/sample.sh** to check whether Kylin is running properly.

11.12 What If Data Failed to Be Exported from MRS to an OBS Encrypted Bucket?

MRS supports OBS bucket encryption from version 1.9.x. If you want to use encrypted OBS buckets, use MRS cluster version 1.9.x or later.

11.13 How Do I Install HSS on MRS Cluster Nodes?

You can install Host Security Service (HSS) on your MRS cluster nodes if the OS versions of the nodes are supported by HSS agents.

Perform the following steps to install HSS:

1. Check whether the OS versions of the MRS cluster nodes are supported by HSS agents.

   To query the OS versions, perform the following operations:

   a. Log in to the MRS management console.

   b. Choose **Clusters** > **Active Clusters**, select a running cluster, and click the cluster name to access its details page.

   c. Click the **Nodes** tab. Click the **master** or **core** node group, and click any node in the node name list to access the cloud service basic information page.

d. Click **Basic Information**, view the image in the ECS information, and confirm the EulerOS version in the image.

   For example, **EulerOS_2.5_x86_64** indicates that the OS version is EulerOS 2.5 (64-bit).

e. Check whether the OS versions of the MRS nodes are supported by HSS agents.

   - If yes, go to **2** to install HSS agents.
   - If no, HSS cannot be installed.

2. Purchase HSS quotas and log in to each node in the MRS cluster to install HSS agents. After the agents are installed, you can enable HSS.

   NOTE

   Ensure you have purchased HSS in your MRS cluster node region and have used the installation package or installation command in the region to install HSS agents on your nodes.

12 Cluster Access

12.1 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.

12.2 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
     i. Choose **Services** > **ZooKeeper**. On the displayed page, click the **Instance** tab to view the business IP address of a ZooKeeper instance.

ii. 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.

12.3 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. For details, see **Purchasing an ECS**.

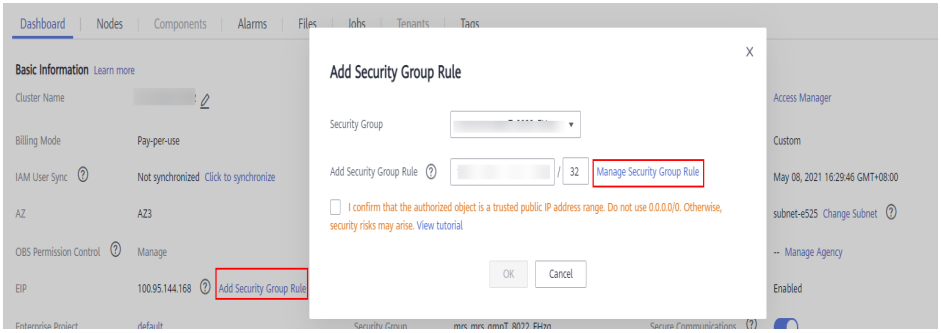
Set **AZ**, **VPC**, and **Security Group** of the ECS to the same values as those of the cluster to be accessed.

Step 2 On the VPC management console, apply for an EIP and bind it to the ECS.

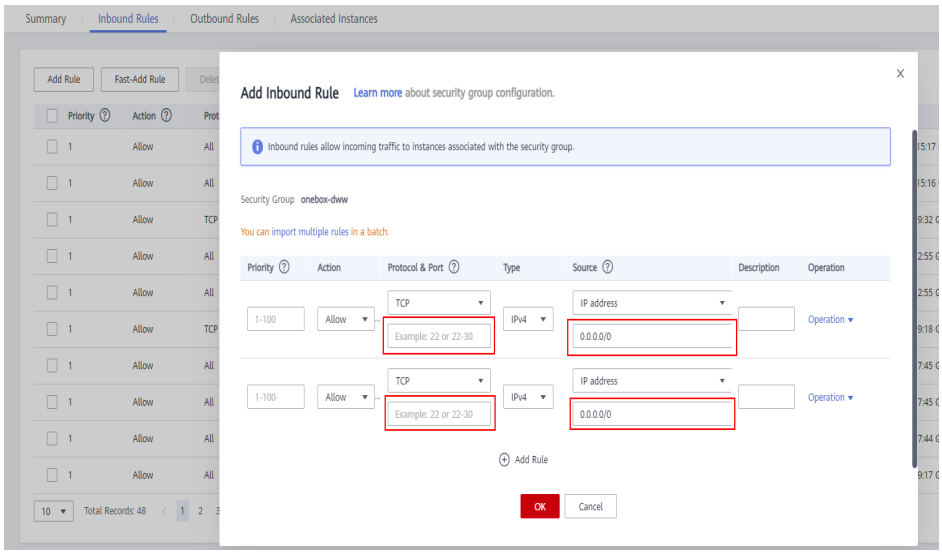
For details, see **Assigning an EIP and Binding It to an ECS**.

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

Creating a Windows ECS Outside the Cluster to Access MRS Manager

Create a Windows ECS on the ECS console to access MRS Manager. For details, see **Accessing MRS Manager Using a Windows ECS**.

13 Big Data Service Development

13.1 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
```

13.2 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.

13.3 Where Are the JAR Files and Environment Variables of Hadoop Stored?

- **hadoopstreaming.jar**: **/opt/share/hadoop-streaming-*** (***** indicates the Hadoop version.)
- JDK environment variables: **/opt/client/JDK/component_env**
- Hadoop environment variables: **/opt/client/HDFS/component_env**
- Hadoop client: **/opt/client/HDFS/hadoop**

13.4 What Compression Algorithms Does HBase Support?

HBase supports the Snappy, LZ4, and gzip compression algorithms.

13.5 Can MRS Write Data to HBase Through the HBase External Table of Hive?

No. Hive on HBase supports only data query.

13.6 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.

13.7 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.
  ```


13.8 How Do I Connect to HBase of MRS Through HappyBase?

ThriftServer1 and ThriftServer2 cannot coexist. The HBase service of the MRS cluster uses ThriftServer2. However, HappyBase can connect to HBase only through the ThriftServer1 interface. Therefore, use Python to directly connect to HBase. For details, see **https://github.com/huaweicloud/huaweicloud-mrs-example/blob/mrs-1.8/src/hbase-examples/hbase-python-example/DemoClient.py**.

13.9 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

13.10 How Do I Change the Number of HDFS Replicas?

1. Go to the HDFS service configuration page.
   - For MRS 1.8.10 or earlier clusters

     Log in to MRS Manager (see **Accessing MRS Manager**), choose **Services** > **HDFS** > **Service Configuration**, and select **All** from the **Basic** drop-down list.
   - For MRS 1.8.10 or later and MRS 2.*x*, click the cluster name. On the cluster details page that is displayed, 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.)

- For MRS cluster version 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.

13.11 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.

For more information, see **List of Open Source Component Ports**.

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

Parameter	Default Port	Port Description
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
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

Parameter	Default Port	Port Description
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
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

Parameter	Default Port	Port Description
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
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

Parameter	Default Port	Port Description
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

13.12 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

13.13 What Is the Recommended Number Type of DynamoDB in Hive Tables?

smallint is recommended.

13.14 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.

13.15 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.

13.16 Where Can I Download the Dependency Package (com.huawei.gaussc10) in the Hive Sample Project?

MRS does not have the **com.huawei.gaussc10** dependency package, which is an GaussDB dependency package and does not need to be configured. Exclude this package when building a Maven project.

13.17 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;
```

13.18 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 13-1 Changing the **component_env** file

  b. Run the following command to verify the configuration:

  source /opt/Bigdata/client/bigdata_env

  beeline -e " use default;show tables;"

13.19 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.

13.20 What If an Excel File Downloaded on Hue Failed to Open?

NOTE

This section applies only to versions earlier than MRS 3.*x*.

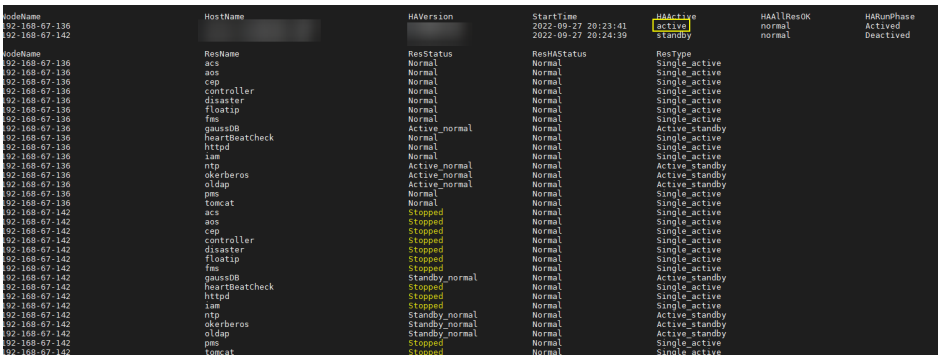
1. Log in to a Master node as user **root** and switch to user **omm**.

   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 13-2 Active OMS node



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* are the same as those in the actual environment.)

ProxyHTMLEnable On
SetEnv PROXY_PREFIX=https://[proxy_ip]:[proxy_port]
ProxyHTMLURLMap (https?:\/\/[^:]*:[0-9]*.*) ${PROXY_PREFIX}/proxyRedirect=$1 RV

Figure 13-3 Content to be deleted

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 13-4 Content to be deleted

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.

13.21 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)**.

13.22 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.

13.23 How Do I Obtain the Client Version of MRS Kafka?

Run the **--bootstrap-server** command to query the information about the client.

13.24 What Access Protocols Are Supported by Kafka?

Kafka supports PLAINTEXT, SSL, SASL_PLAINTEXT, and SASL_SSL.

13.25 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.

13.26 What Compression Algorithms Does Kudu Support?

Kudu supports **Snappy**, **LZ4**, and **zlib**. **LZ4** is used by default.

13.27 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.

13.28 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 00000000000000000000000: 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.

13.29 What Are the Differences Between Sample Project Building and Application Development? Is Python Code Supported?

- The sample project and application development in MRS are the same. You can select either of them.
- MRS supports Python code.

13.30 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.

13.31 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.

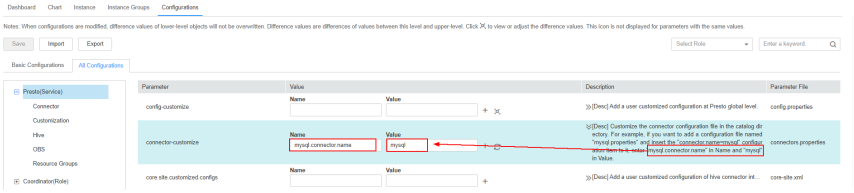
c. Click the **Components** tab and then **Presto** in the component list. On the page that is displayed, click the **Configurations** tab then the **All Configurations** sub-tab.

d. On the Presto configuration page that is displayed, find **connector-customize**.

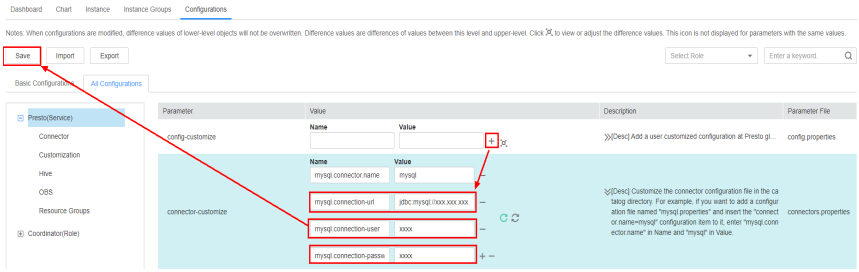
e. Set **Name** and **Value** as follows:

Name: **mysql.connector.name**

Value: **mysql**



f. Click the plus sign (+) to add three more fields and set **Name** and **Value** according to the table below. Then click **Save**.



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://15[redacted]20
       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.

13.32 How Do I Update the Ranger Certificate?

MRS 1.9.3 is used as an example. Replace it with the actual cluster version. After the certificate is updated, manually clear the alarm indicating that the certificate file is invalid or about to expire.

NOTE

After the Ranger certificate is updated, its validity period is 10 years.

After the Ranger certificate expires, the Ranger web UI is still accessible, but a message indicating that the certificate is untrusted will be displayed when you access the web UI.

- If Ranger is not installed in the cluster, log in to each master node and run the following command to rename the certificate file:

  mv /opt/Bigdata/MRS_1.9.3/install/MRS-Ranger-1.0.1/ranger/ranger-1.0.1-admin/ranger-admin-keystore.jks /opt/Bigdata/MRS_1.9.3/install/MRS-Ranger-1.0.1/ranger/ranger-1.0.1-admin/ranger-admin-keystore.jks_bak

- If Ranger has been installed in the cluster, update the certificate as follows:

a. Download **MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz** from the obs-patch bucket and upload it to the **/tmp** directory on the node where the active RangerAdmin instance of the cluster runs.

   On MRS Manager, choose **Service** > **Ranger** > **Instance** and obtain the IP address of the node where the active RangerAdmin instance runs.

   - CN-Hong Kong: **https://mrs-patch-ap-southeast-1.obs.ap-southeast-1.myhuaweicloud.com/MRS_Common_Script/MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz**
   - AP-Bangkok: **https://mrs-patch-ap-southeast-2.obs.ap-southeast-2.myhuaweicloud.com/MRS_Common_Script/MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz**
   - AP-Singapore: **https://mrs-patch-ap-southeast-3.obs.ap-southeast-3.myhuaweicloud.com/MRS_Common_Script/MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz**
   - LA-Sao Paulo: **https://mrs-container1-patch-sa-brazil-1.obs.myhuaweicloud.com/MRS_Common_Script/MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz**
   - LA-Mexico City: **https://mrs-container1-patch-na-mexico-1.obs.myhuaweicloud.com/MRS_Common_Script/MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz**

b. Log in to the node where the active RangerAdmin instance is located and run the following commands:

   cd /tmp

   chmod 700 MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz

   chown omm:wheel MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz

   su - omm

   cd /tmp

   tar -zxvf MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz

c. Replace the certificate files.

   cd updateRangerJks

   sh updateRangerJks.sh $***{IP address of the active master node}* **$*{IP address of the active RangerAdmin node}* **$***{Certificate password}*

   NOTE

   - This script will restart the controller process. During the restart process, the MRS Manager page may not be viewed.
   - Obtain the IP address of the active master node from **Hosts** on MRS Manager.
   - To obtain the IP address of the active RangerAdmin node, choose **Services** > **Ranger** > **Instances** on MRS Manager.
   - *${Certificate password}* is a user-defined password.

d. Log in to the MRS console.

e. Choose **Clusters > Active Clusters** and click a cluster name to go to the cluster details page.

f. Choose **Components** > **Ranger** > **Service Configuration** and modify the RangerAdmin configuration.

   i. Search for the **policymgr_https_keystore_password** and change its value to the certificate password entered in **c**, that is, *${Certificate password}*.

      You are advised to copy and paste the password. If the passwords are different, Ranger will fail to restart.

   ii. Save the configuration and perform a rolling restart of RangerAdmin.

g. Verify that you can log in to the RangerAdmin web UI.

   i. Choose **Components** > **Ranger** > **Service Status**. In **Ranger Summary**, click **RangerAdmin** corresponding to **Ranger Web UI**.

   ii. On the Ranger web UI login page, the default username for MRS cluster 1.9.2 is **admin** and the password is **admin@12345**. The default username for MRS cluster 1.9.3 or later is **admin** and the password is **ranger@A1!**.

      After logging in to the Ranger Web UI for the first time, change the password and keep it secure.

h. Log in to the node where the RangerAdmin instance is located and delete the temporary files.

   rm -rf /tmp/updateRangerJks

   rm -rf /tmp/updateRangerJks.tar.gz

   For a cluster with a custom topology, if the active master and RangerAdmin instances are not on the same node, log in to the active master node and delete temporary files.

13.33 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 *Client installation directory*/**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

13.34 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 *Client installation directory*/**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.

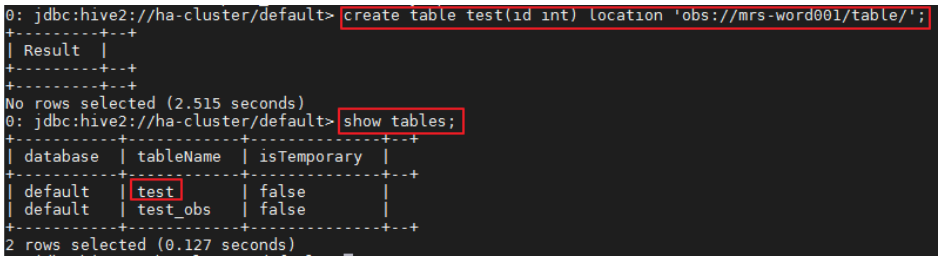
   create table test(id int) location 'obs://mrs-word001/table/';

6. Query all tables.

   show tables;

   If the table **test** is displayed in the command output, OBS is successfully accessed.

   Figure 13-5 Returned table name



7. Press **Ctrl+C** to exit the Spark Beeline.

13.35 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.

13.36 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.

13.37 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:Click the name of the ResourceManager node, and choose **More** > **Synchronize Configuration**.In the dialog box displayed, deselect **Restart services or instances whose configurations have expired** and click **Yes**.

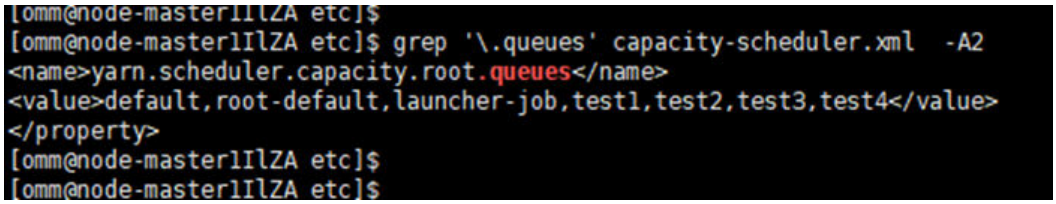
Step 3 Click **Yes** to synchronize the configuration.

Step 4 Log in to the Master nodes as user **root**.

Step 5 Run the **cd /opt/Bigdata/MRS_Current/*_*_ResourceManager/etc_UPDATED/** command to go to the **etc_UPDATED** directory.

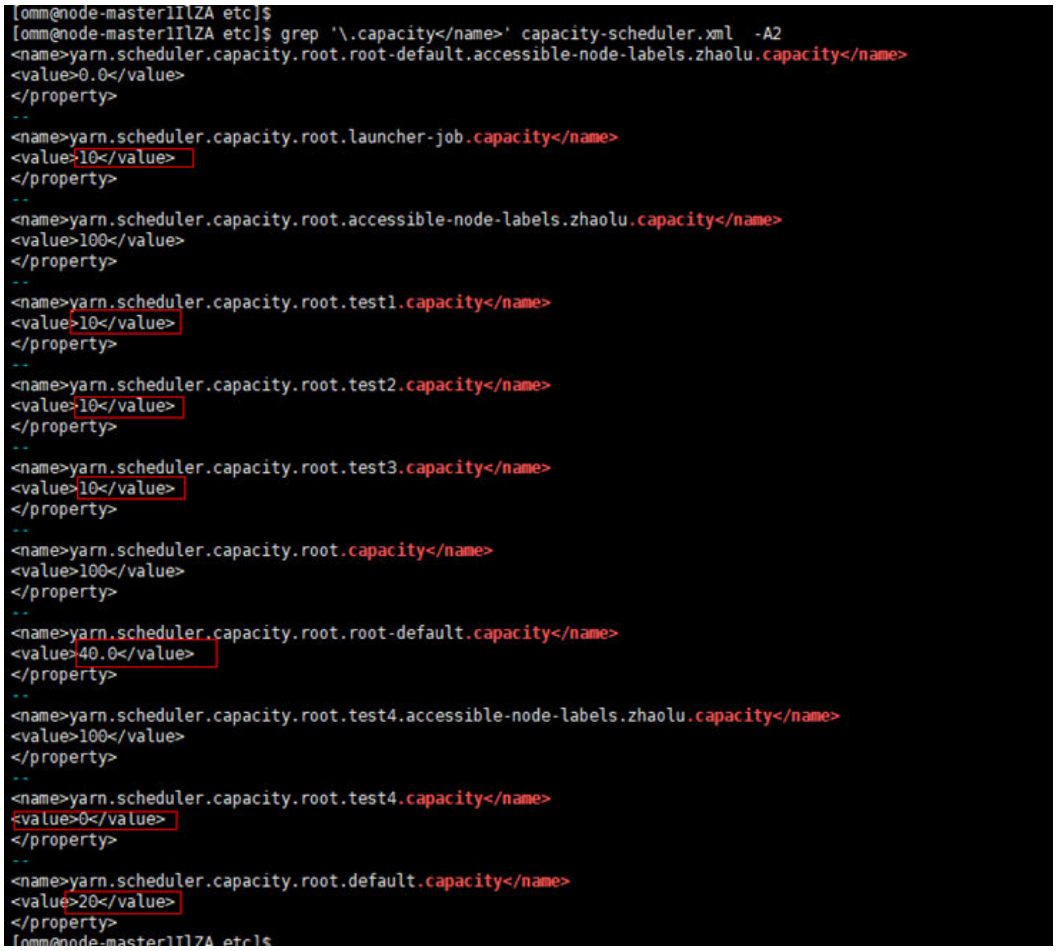
Step 6 Run the **grep '\.queues' capacity-scheduler.xml -A2** command to display all configured queues and check whether the queues are consistent with those displayed on Manager.

root-default is hidden on the Manager page.



Step 7 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.



Step 8 Log in to MRS Manager, and select **Hosts**.

Step 9 Determine the active Master node. The host name of the active Master node starts with a solid pentagon.

Step 10 Log in to the active Master node as user **root**.

Step 11 Run the **su - omm** command to switch to user **omm**.

Step 12 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 13 Repeat **Step 1** to **Step 7** 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

13.38 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:

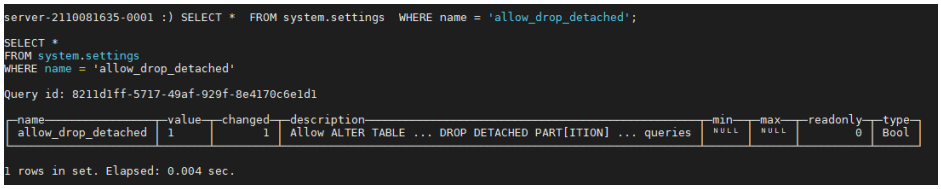
```
[root@server-2110082001-0017 hadoopclient]# clickhouse client --host 192.168.42.90 --secure -m
ClickHouse client version 21.3.4.25.
Connecting to 192.168.42.90:21427.
Connected to ClickHouse server version 21.3.4 revision 54447.
```

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';



Step 7 Run the **q;** command to exit the ClickHouse client.

----End

13.39 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.

13.40 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.

13.41 How Do I Obtain a Spark JAR File?

Huawei provides Huawei Mirrors for you to download all dependency JAR files of sample projects, but you need to download the open source dependency JAR files from the Maven central repository or other custom repositories.

NOTE

Ensure the following conditions are met before you use a development tool in a local environment to download dependency JAR files:

- The local network is normal.

  Open a browser and enter the Huawei Mirrors URL in the address box to check whether your access is normal. If the access is abnormal, check whether your local network is accessible.

- Check whether the proxy is enabled for the development tool. If it is enabled, disable it by performing the following operations.

  Take IntelliJ IDEA 2020.2 as an example. Choose **File** > **Settings** > **Appearance & Behavior** > **System Settings** > **HTTP Proxy**, select **No proxy**, and click **OK** to save the configuration.

13.42 Why Is an Alarm Generated When the NameNode Process Is Not Restarted After the hdfs-site.xml File Is Modified?

If you do not restart the NameNode process after modifying the **dfs.namenode.checkpoint.period** parameter, an alarm may be falsely reported. In this case, you need to restart the NameNode process as soon as possible.

13.43 It Takes a Long Time for Spark SQL to Access Hive Partitioned Tables Before Job Startup

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 API

14.1 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**.

15 Cluster Management

15.1 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

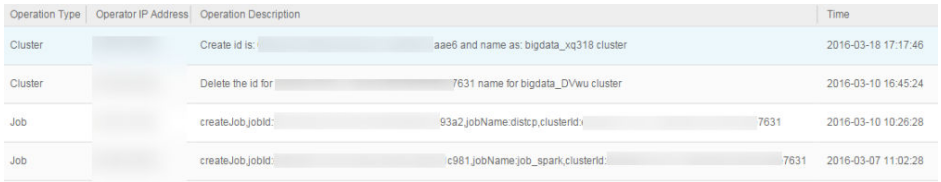
15.2 How Do I View Log Information?

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 15-1 shows the operation logs.

Figure 15-1 Log information



Operation Type	Operator IP Address	Operation Description	Time
Cluster		Create id is: aae6 and name as: bigdata_xq318 cluster	2016-03-18 17:17:46
Cluster		Delete the id for 7631 name for bigdata_DVwu cluster	2016-03-10 16:45:24
Job		createJob,jobId: 93a2,jobName:distcp,clusterId: 7631	2016-03-10 10:26:28
Job		createJob,jobId: c981,jobName:job_spark,clusterId: 7631	2016-03-07 11:02:28

15.3 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.

15.4 How Do I Add Services to an MRS Cluster?

You cannot install new components for a created cluster whose version is MRS 3.1.0 or earlier. To install streaming cluster components, you need to create a streaming cluster or hybrid cluster and install required components.

You can add components to MRS 3.1.2-LTS.3 or later custom clusters. For details, see **Managing Services**.

15.5 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.

You can add components to MRS 3.1.2-LTS.3 or later custom clusters. For details, see **Managing Services**.

15.6 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**.

15.7 Do I Need to Shut Down a Master Node Before Upgrading Its Specifications?

The Master node automatically shuts down during the specifications upgrade, and automatically starts after the upgrade. No manual intervention is required.

15.8 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. For details, see **Expanding Capacity for an In-use EVS Disk**. MRS nodes are installed using public images and support the capacity expansion of in-use EVS disks.

15.9 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.

15.10 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.

15.11 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.

15.12 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**.

15.13 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.

15.14 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 15-2 knox memory

----End

15.15 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.

Table 15-1 Python versions supported by MRS

MRS Cluster Version	Python Version
MRS 3.1.0	Python 3.8.0
MRS 3.0.5	Python 3.7.0
MRS 3.0.2	Python 3.7.0
MRS 2.1.1	Python 3.6.8
MRS 2.1.0	Python 3.6.8
MRS 1.9.3	Python 3.6.8

15.16 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*

15.17 How Do I Upload a Local File to a Node Inside a 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 the target cluster to go to its details page.

Step 3 On the **Nodes** page, click a node name to log in to the ECS console.

Step 4 Bind an elastic IP address to the cluster node. For details, see **Assigning an EIP and Binding It to an ECS**.

Step 5 Upload the local file to the cluster node. For details, see **How Can I Upload a File to an ECS?**

----End

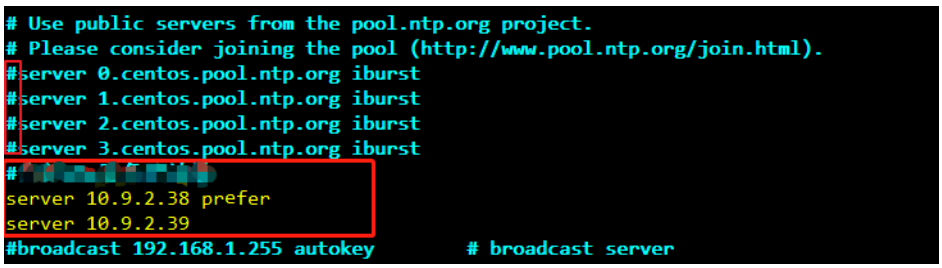
15.18 How Do I Do If the Time on MRS Nodes Is Incorrect?

- 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 15-3 Adding the master node IP addresses



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:

15.19 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
```

15.20 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 ~]$ 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:uChnRUbhh1HYxpfOZibS0zymlKXMlaFyvn0IMpiZjg /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.

15.21 How Do I Adjust the Memory Size of the manager-executor Process?

Symptom

The Huawei-developed **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

   b. Run the following command to restart the process:

      sh /opt/executor/bin/shutdown.shsh /opt/executor/bin/startup.sh

15.22 Can I Modify a Master Node in an Existing MRS Cluster?

Yes.

You can scale up the specifications of a master node in an existing MRS cluster containing at least two master nodes. For details, see **Scaling Up Master Node Specifications**.

16 Kerberos Usage

16.1 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.

16.2 What Are the Ports of the Kerberos Authentication Service?

The Kerberos authentication service uses ports 21730 (TCP), 21731 (TCP/UDP), and 21732 (TCP/UDP).

16.3 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.

16.4 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.

16.5 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

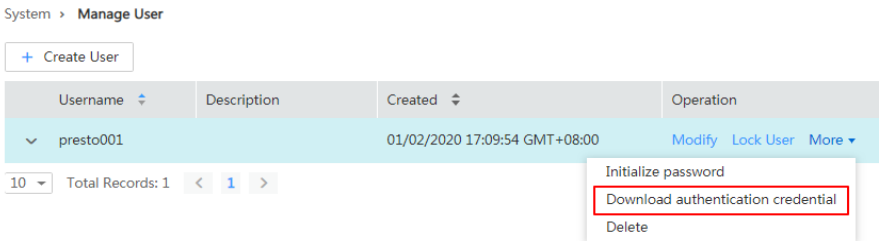
3. Access Presto in a cluster with Kerberos authentication enabled.
   a. Log in to MRS Manager and create a role with the **Hive Admin Privilege** permission, for example, **prestorerole**.
   b. Create a user, for example, **presto001**, who belongs to the **Presto** and **Hive** groups, and bind the 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 and choose **System** > **Manage User**. Locate the row containing the new user, click **More**, and select **Download authentication credential**.

        Figure 16-1 Downloading the Presto user authentication credential

      - 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 16-2 Return result

j. Run **exit** to exit the client.

16.6 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

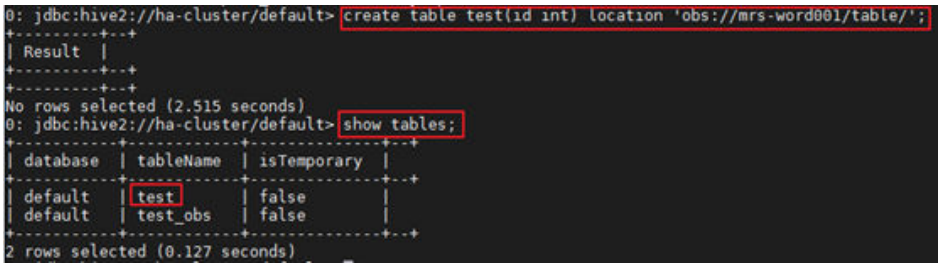
5. Run commands on Spark Beeline. For example, create the table **test** in the **obs://mrs-word001/table/** directory.

   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 16-3 Returned table name



7. Press **Ctrl+C** to exit Spark Beeline.

16.7 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**

17 Metadata Management

17.1 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.

  NOTE

  To query the IP address of the active DBServer node, log in to FusionInsight Manager, choose **Cluster** > **Services** > **DBService**, and click the **Instance** tab.

- 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.

