Copyright © Huawei Technologies Co., Ltd. 2020. All rights reserved.
No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of 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 Account Passwords .................................................................................................................. 1
  1.1 Resetting or Changing the Password of User admin.......................................................... 1

2 Account Permissions ............................................................................................................... 2
  2.1 When a User Uses the AK/SK to Call the MRS Cluster Host List Interface, the Message "User do not have right to access cluster" Is Displayed.............................................................................................. 2

3 Accessing the Web Pages ...................................................................................................... 4
  3.1 Failed to Access MRS Manager........................................................................................................ 4
  3.2 Accessing MRS Manager Using Mac.............................................................................................. 5
  3.3 Failed to Log In to MRS Manager After the Python Upgrade....................................................... 8

4 APIs ............................................................................................................................................. 9
  4.1 Failed to Call an API to Create a Cluster......................................................................................... 9

5 Cluster Management ............................................................................................................. 10
  5.1 Failed to Scale In Task Nodes........................................................................................................ 10
  5.2 Expired OBS Certificate in a Cluster........................................................................................... 11
  5.3 Adding a New Disk to an MRS Cluster............................................................................................ 13
  5.4 Replacing a Disk in an MRS Cluster............................................................................................. 16
  5.5 MRS Backup Failure.................................................................................................................... 19
  5.6 Inconsistency Between df and du Command Output on the Core Node........................................ 20
  5.7 Disassociating a Subnet from the ACL Network.......................................................................... 20
  5.8 MRS Becomes Abnormal After hostname Modification............................................................. 21
  5.9 Determining Whether a Process Is Killed...................................................................................... 22
  5.10 Failed to Configure Cross-Cluster Mutual Trust...................................................................... 24
  5.11 Network Is Unreachable When Using pip3 to Install the Python Package in an MRS Cluster...... 25
  5.12 Connecting the Open-Source confluent-kafka-go to the Security Cluster of MRS.................. 26
  5.13 Failed to Periodically Back Up an MRS 1.7.2 Cluster................................................................. 28
  5.14 Failed to Download the MRS Cluster Client.............................................................................. 28
  5.15 An Error Is Reported When a Flink Job Is Submitted in a Cluster with Kerberos Authentication Enabled............................................................................................................................ 29
  5.16 Scale-out Failure......................................................................................................................... 31
  5.17 Error Occurs When MRS Executes the Insert Command Using Beeline.................................. 33

6 Using HBase ............................................................................................................................ 34
  6.1 Slow Response to HBase Connection......................................................................................... 34
6.1 Failed to Authenticate the HBase User
6.2 RegionServer Start Failure Due to Port Occupation
6.3 HBase Startup Failure Due to Insufficient Memory on the Node
6.4 HBase Startup Failure Due to Insufficient Memory on the Node
6.5 HBase Service Unavailable Due to Poor HDFS Performance
6.6 HBase Startup Failure Due to Inappropriate Parameter Settings
6.7 RegionServer Start Failure Due to Residual Processes
6.8 HBase Startup Failure Caused by Quota Settings on HDFS
6.9 HBase Startup Failure Caused by Initial Timeout Reservation
6.10 HBase Startup Failure Caused by Corrupted Version Files
6.11 High CPU Usage of RegionServer Even No Services Are Running
6.12 HBase Startup Failure and the "FileNotFoundException" Error Is Recorded in RegionServer Logs
6.13 The Number of RegionServers Displayed on the Native Page Is Greater Than the Actual Number After HBase Is Started
6.14 The RegionServer Instance Is Abnormal and in the Concerning State
6.15 HBase Startup Failure in the Newly Installed Cluster
6.16 HBase Startup Failure Due to the Lost ACL Table Directory
6.17 HBase Fails to Start After the Cluster Is Powered Off and On
6.18 HBase Is Abnormal

7 Using Hive
7.1 Content Recorded in Hive Logs
7.2 Causes of Hive Startup Failure
7.3 "Cannot modify xxx at runtime" Is Reported When the set Command Is Executed in a Security Cluster
7.4 How to Specify a Queue When Hive Submits a Task
7.5 How to Set Map and Reduce Memory on the Client
7.6 How to Specify the Output File Compression Format When Importing a Table
7.7 The desc Table Is Too Long to Be Completely Displayed
7.8 Data Inserted After the Partition Column Is Added Is Displayed as NULL
7.9 A Newly Created User Has No Query Permissions
7.10 An Error Is Reported When Executing SQL to Submit a Task to a Specified Queue
7.11 An Error Is Reported When The load data inpath Command Is Executed
7.12 An Error Is Reported When The load data local inpath Command Is Executed
7.13 An Error Is Reported When The create external table Command Is Executed
7.14 An Error Is Reported When The dfs -put Command Is Executed On The Beeline Client
7.15 An Error Is Reported Indicating Insufficient Permissions When the set role admin Command Is Executed
7.16 An Error Is Reported When a UDF Is Created Using Beeline
7.17 Hive Status Is Bad
7.18 Hive Service Status Is Partially Healthy
7.19 Difference Between Hive Service Health Status and Hive Instance Health Status
7.20 Alarms in Hive and Triggering Scenarios
7.21 "authentication failed" Is Displayed During Shell Client Connection
7.22 Client Displays a Message Indicating ZooKeeper Access Failure
9.2 How Do I Use Python3.x to Connect to Kafka in a Security Cluster?............................................................... 109
9.3 Flume Normally Connects to Kafka But Fails to Send Messages................................................................................. 111
9.4 Producer Fails to Send Data and "NullPointerException" Is Thrown .......................................................................... 114
9.5 Producer Fails to Send Data and Throws "TOPIC_AUTHORIZATION_FAILED"......................................................... 117
9.6 Producer Occasionally Fails to Send Data and the Log Displays "Too many open files in system"............................ 119
9.7 Consumer Is Initialized Successfully, But the Specified Topic Message Cannot Be Obtained from Kafka...................... 122
9.8 Consumer Fails to Consume Data and Remains in the Waiting State............................................................................. 127
9.9 SparkStreaming Fails to Consume Kafka Messages, and the Message "Error getting partition metadata" Is Displayed.............................................................. 130
9.10 Consumer Fails to Consume Data in a Newly Created Cluster, and the Message "GROUP_COORDINATOR_NOTAVAILABLE" Is Displayed................................................................................................................... 132
9.11 SparkStreaming Fails to Consume Kafka Messages, and the Message "Couldn't find leader offsets" Is Displayed........................................................................................................... 133
9.12 Consumer Fails to Consume Data and the Message " SchemaException: Error reading field 'brokers'" Is Displayed................................................................................................................................................. 135
9.13 Checking Whether Data Consumed by a Customer Is Lost..................................................................................... 136
9.14 Failed to Start a Component Due to Account Lock.................................................................................................. 137
9.15 Kafka Broker Reports Abnormal Processes and the Log Shows "IllegalArgumentException"......................... 138
9.16 Kafka Topics Cannot Be Deleted............................................................................................................................ 139
9.17 Error "AdminOperationException" Is Displayed When a Kafka Topic Is Deleted..................................................... 143
9.18 When a Kafka Topic Fails to Be Created, "NoAuthException" Is Displayed............................................................... 144
9.19 Failed to Set an ACL for a Kafka Topic, and "NoAuthException" Is Displayed............................................................. 146
9.20 When a Kafka Topic Fails to Be Created, "NoNode for /brokers/ids" Is Displayed.................................................... 148
9.21 When a Kafka Topic Fails to Be Created, "replication factor larger than available brokers" Is Displayed.................. 150
9.22 Consumer Repeatedly Consumes Data.................................................................................................................... 152
9.23 Partitions Are Unevenly Distributed on Nodes During the Creation of Kafka Topics.............................................. 154
9.24 Leader for the Created Kafka Topic Partition Is Displayed as none......................................................................... 154
9.25 Safety Instructions on Using Kafka.......................................................................................................................... 157
9.26 Obtaining Kafka Consumer Offset Information........................................................................................................... 162
9.27 Adding or Deleting Configurations for a Topic............................................................................................................. 164
9.28 Reading the Content of the _consumer_offsets Internal Topic.................................................................................. 165
9.29 Configuring Logs for Shell Commands on the Client.................................................................................................. 166
9.30 Obtaining Topic Distribution Information..................................................................................................................... 167
9.31 Kafka HA Usage Description.......................................................................................................................................... 168
9.32 A Kafka Cluster Cannot Be Operated Using the Kafka Shell Command........................................................................... 171
9.33 Kafka Producer Writes Oversized Records.................................................................................................................... 172
9.34 Kafka Consumer Reads Oversized Records.................................................................................................................. 173

10 Using Storm.................................................................................................................................................................... 175
10.1 Invalid Hyperlink of events on the Storm UI.................................................................................................................. 175
10.2 Failed to Submit a Topology................................................................................................................................................. 176
10.3 Topology Submission Fails and the Message "Failed to check principle for keytab" Is Displayed..................................... 179
10.4 The Worker Log Is Empty After a Topology Is Submitted............................................................................................. 180
10.5 Worker Runs Abnormally After a Topology Is Submitted and Error "Failed to bind to:host:ip" Is Displayed.................................................................................................................................................................................. 183
10.6 "well-known file is not secure" Is Displayed When the jstack Command Is Used to Check the Process Stack.................................................................................................................................................................................. 185
10.7 When the Storm-JDBC plug-in is used to develop Oracle write Bolts, data cannot be written into the Bolts........................................................................................................................................................................................................ 188
10.8 The GC Parameter Configured for the Service Topology Does Not Take Effect.................................................. 190
10.9 Internal Server Error Is Displayed When the User Queries Information on the UI.................................................. 191

11 Using Flume ............................................................................................................................................................................. 194
11.1 Class Cannot Be Found After Flume Submits Jobs to Spark Streaming................................................................. 194
11.2 Failed to Install a Flume Client........................................................................................................................................ 195
11.3 A Flume Client Cannot Connect to the Server................................................................................................................. 195
11.4 Flume Data Fails to Be Written to the Component........................................................................................................ 196
11.5 Flume Server Process Fault............................................................................................................................................. 197
11.6 Flume Data Collection Is Slow......................................................................................................................................... 197
11.7 Flume Startup Failure...................................................................................................................................................... 198

12 Using Hue ...................................................................................................................................................................................... 200
12.1 A Job Is Running on Hue.................................................................................................................................................. 200
12.2 HQL Fails to Be Executed on Hue Using Internet Explorer......................................................................................... 201
12.3 Hue (Active) Cannot Open Web Pages........................................................................................................................... 201
12.4 Failed to Access the Hue Web UI................................................................................................................................ 202
12.5 HBase Tables Cannot Be Loaded on the Hue Web UI........................................................................................................ 202

13 Using HDFS ................................................................................................................................................................................... 204
13.1 All NameNodes Become the Standby State After the NameNode RPC Port of HDFS Is Changed............. 204
13.2 An Error Is Reported When the HDFS Client Is Used After the Host Is Connected Using a Public Network IP Address..................................................................................................................................................... 205
13.3 Failed to Use Python to Remotely Connect to the Port of HDFS.................................................................................. 206
13.4 HDFS Capacity Usage Reaches 100%, Causing Unavailable Upper-layer Services Such as HBase and Spark.................................................................................................................................................................. 207
13.5 An Error Is Reported During HDFS and Yarn Startup............................................................................................... 208
13.6 HDFS Permission Setting Error........................................................................................................................................ 209
13.7 A DataNode of HDFS Is Always in the Decommissioning State.................................................................................... 210
13.8 HDFS Startup Failure Due to Insufficient Memory...................................................................................................... 211
13.9 A Large Number of Blocks Are Lost in HDFS due to the Time Change Using ntpdate............................ 213
13.10 CPU Usage of a DataNode Reaches 100% Occasionally, Causing Node Loss (SSH Connection Is Slow or Fails).................................................................................................................................................. 215
13.11 Manually Performing Checkpoints When a NameNode Is Faulty for a Long Time.................................................. 216
13.12 Common File Read/Write Faults................................................................................................................................... 217
13.13 Maximum Number of File Handles Is Set to a Too Small Value, Causing File Reading and Writing Exceptions.................................................................................................................................................. 218
13.14 A Client File Fails to Be Closed After Data Writing....................................................................................................... 219
13.15 File Fails to Be Uploaded to HDFS Due to File Errors..................................................................................................... 222
13.16 After dfs.blocksize Is Configured on the Interface and Data Is Put, the Original Block Size Keeps Unchanged.................................................................222
13.17 Failed to Read Files, and "FileNotFoundException" Is Displayed...............................................................223
13.18 Failed to Write Files to HDFS, and "item limit of / is exceeded" Is Displayed..............................................224
13.19 Adjusting the Log Level of the Shell Client......................................................................................................224
13.20 File Read Fails, and "No common protection layer" Is Displayed.....................................................................225
13.21 Failed to Write Files Because the HDFS Directory Quota Is Insufficient.......................................................226
13.22 Balancing Fails, and "Source and target differ in block-size" Is Displayed......................................................227
13.23 A File Fails to Be Queried or Deleted, and the File Can Be Viewed in the Parent Directory (Invisible Characters)..................................................................................................................228
13.24 Uneven Data Distribution Due to Non-HDFS Data Residuals........................................................................229
13.25 Uneven Data Distribution Due to the Client Installation on the DataNode......................................................230
13.26 Handling Unbalanced DataNode Disk Usage on Nodes..................................................................................230
13.27 Locating Common Balance Problems (Updated on July 27, 2019)...............................................................231
13.28 HDFS Displays Insufficient Disk Space But 10% Disk Space Remains.........................................................233
13.29 An Error Is Reported When the HDFS Client Is Installed on the Core Node in a Common Cluster.233

14 Using Yarn..................................................................................................................................................................235
14.1 Plenty of Jobs Are Found After Yarn Is Started..................................................................................................235
14.2 Error Message "GC overhead" Is Displayed on the Client During Task Submission by Using the Hadoop Jar Command..................................................................................................................237
14.3 Disk Space Is Used Up Due to Oversized Aggregated Logs of Yarn...............................................................238

15 Using DBService......................................................................................................................................................240
15.1 DBServer Instance Status Abnormal..................................................................................................................240
15.2 DBServer Instance Is Always in the Concerning State.......................................................................................241
15.3 Default Port Number 20050 or 20051 Is Occupied...............................................................................................242
15.4 DBServer Instance Is Always in the Concerning State Because the /tmp Directory Permission Is Incorrect..................................................................................................................................................243
15.5 DBService Backup Failure....................................................................................................................................244
15.6 DBService Is Normal But Components Cannot Connect to It............................................................................245
15.7 DBServer Startup Failure.....................................................................................................................................246
15.8 DBService Backup Failure Due to Unreachable Floating IP Address...............................................................247
15.9 Startup Failure Due to Lost DBService Configuration File................................................................................248

16 Using Flink................................................................................................................................................................251
16.1 "IllegalConfigurationException: Error while parsing YAML configuration file: "security.kerberos.login.keytab" Is Displayed When a Command Is Executed on an Installed Client......251
16.2 "IllegalConfigurationException: Error while parsing YAML configuration file" Is Displayed When a Command Is Executed After Configurations of the Installed Client Are Changed .......................................................252
16.3 The yarn-session.sh Command Fails to Be Executed When the Flink Cluster Is Created................................253
16.4 Failed to Create a Cluster by Executing the yarn-session Command When a Different User Is Used ..........254
16.5 Flink Service Program Fails to Read Files on the NFS Disk............................................................................255
16.6 New Topic 6..........................................................................................................................................................256

17 Using Impala..............................................................................................................................................................257
17.1 Failed to Connect to impala-shell.............................................................................................................................. 257

18 Using Presto........................................................................................................................................................................ 259
18.1 During sql-standard-with-group Configuration, a Schema Fails to Be Created and the Error Message
"Access Denied" Is Displayed.................................................................................................................................................259

19 Using Alluixo........................................................................................................................................................................... 261
19.1 Error Message "Does not contain a valid host:port authority" Is Reported When Alluixo Is in HA
Mode.................................................................................................................................................................................... 261

20 Using ZooKeeper................................................................................................................................................................... 262
20.1 Accessing ZooKeeper from an MRS Cluster.................................................................................................................... 262

21 Accessing OBS...................................................................................................................................................................... 263
21.1 When Using the MRS Multi-user Access to OBS Function, a User Does Not Have the Permission to
Access the /tmp Directory...................................................................................................................................................... 263
21.2 When the Hadoop Client Is Used to Delete Data from OBS, It Does Not Have the Permission for
the .Trash Directory................................................................................................................................................................. 264
1 Account Passwords

1.1 Resetting or Changing the Password of User admin

- If you know the password, change the password following the instructions in Changing the Password of User admin.
- If you forget the password, reset the password following the instructions in Changing the Password of a Component Running User.
2 Account Permissions

2.1 When a User Uses the AK/SK to Call the MRS Cluster Host List Interface, the Message "User do not have right to access cluster" Is Displayed

Issue

When a user uses the AK/SK to call the MRS cluster host list interface /v1.1/<project_id>/clusters/<cluster_id>/hosts, the message "User do not have right to access cluster" is displayed.

Cause Analysis

When the user uses the AK/SK to obtain the host list of the MRS cluster, parameters such as project_id are not filled in the request header. As a result, the project_id of the token parsed by the cloud service is inconsistent with that of the cluster, and a message is displayed indicating that the user does not have the permission.

Procedure

Before calling interfaces in AK/SK-based authentication mode, collect required information by referring to Table 2-1. For details about how to sign requests and use SDKs, see API Request Signing Guide.

Before constructing an API request, obtain the following information, including the endpoint and URI of the request URL, AK/SK used for signature and authentication, and project ID used to distinguish tenants.

Table 2-1 Information to be collected

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Endpoint of a cloud service in a region. For details, see Regions and Endpoints.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project_ID</strong></td>
<td>Project ID, which needs to be configured in the URI of most APIs to identify different projects.</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>AK/SK</strong></td>
<td>Access key pair, including the Access key ID (AK) and secret access key (SK), which are used to sign API requests.</td>
</tr>
<tr>
<td><strong>URI</strong></td>
<td>API request path and parameters. For details, see <a href="#">API Overview</a>.</td>
</tr>
</tbody>
</table>
| **X-Domain-Id** | Account ID, which is used for:  
  - Obtaining a token for token-based authentication  
  - Calling APIs of global services using AK/SK-based authentication. **X-Domain-Id** needs to be configured in the request header. |
| **X-Project-Id** | Sub-project ID, which is used in multi-project scenarios. To access resources in a sub-project through AK/SK-based authentication, the **X-Project-Id** field must be added to the request header. |

**NOTE**

For details about how to obtain the parameters in Table 2-1, see [API Request Signing Guide](#).
3 Accessing the Web Pages

3.1 Failed to Access MRS Manager

Symptom

The MRS Manager is not accessible after a cluster is created.

Possible Cause

- MRS can be accessed from an external network only after an EIP is bound to an MRS node.
- Port 9022 is disabled. Add a security group rule to enable the port.
- The cluster version must be 1.8.0 or later.

Procedure

Step 1 Log in to the MRS management console, locate the cluster to be accessed in the active cluster list, and click the cluster name.

Step 2 On the node information page, click the name of the node to be accessed, and choose EIPs > Bind EIP.

Step 3 On the Bind EIP page, select a NIC from the Select NIC drop-down list, select an EIP from the Select EIP list, and click OK.

Step 4 After the EIP is bound, enable port 9022 in a security group rule.

   Click the Security Groups tab. Then, click Change Security Group.

   You can select an existing security group, or click Create Security Group to add a security group rule to enable port 9022 for accessing through the public IP address. For details, see Configuring Security Group Rules.

Step 5 After the EIP is added, you can access MRS through https://Elastic IP address:9022/mrsmanager/. If the fault still persists, contact technical support for assistance.
### 3.2 Accessing MRS Manager Using Mac

**Issue**

How do I use Google Chrome to access MRS Manager on a Mac?

**Symptom**

Google Chrome cannot be used to access MRS Manager on a Mac.

**Cause Analysis**

The Mac cannot be accessed due to certificate restrictions.

**Procedure**

1. **Step 1** Copy the MRS Manager login link and use Safari to access MRS Manager.
2. **Step 2** Click visit this website on the Your Connection Is Not Private page.

![Figure 3-1 Accessing MRS Manager using Safari](image)

**Step 3** In the displayed dialog box, click Visit Website.
Step 4  Enter the username and password of your computer to access MRS Manager.

Figure 3-3 Entering the username and password
Step 5 The MRS Manager login page is displayed. At this time, you cannot log in to MRS Manager using Google Chrome. Go to the next step.

Step 6 Among Mac applications, select Keychain Access.

Figure 3-4 Keychain Access

Step 7 In the Keychains area, select login. In the Category area, select Certificates.

Step 8 Locate the certificate issued by console.hwclouds.com, double-click the certificate, and expand Trust.

Step 9 Set When using this certificate to Always Trust and close the dialog box.

Figure 3-5 Trust

Step 10 In the displayed dialog box, enter the login password. Then, you can access MRS Manager using Chrome.

----End
### 3.3 Failed to Log In to MRS Manager After the Python Upgrade

**Issue**

Failed to log in to MRS Manager after Python is upgraded.

**Symptom**

After Python is upgraded, MRS Manager fails to be accessed using the *admin* account and the correct password.

**Possible Cause**

When upgrading Python to Python 3.x, the user modifies the file directory permission of *openssl*. As a result, the LdapServer service cannot be started, causing a login authentication failure.

**Procedure**

1. **Step 1** Log in to the Master node in the cluster as user *root*.
2. **Step 2** Run the `chmod 755 /usr/bin/openssl` command to modify the file directory permission of `/usr/bin/openssl` to *755*.
3. **Step 3** Run the `su omm` command to switch to user *omm*.
4. **Step 4** Run the `openssl` command to check whether the *openssl* mode can be entered.
   - If it can be entered, the permission has been modified successfully. If it cannot be entered, the permission fails to be modified.
   - If the permission fails to be modified, check whether the command is correct or contact O&M personnel.
5. **Step 5** After the permission is modified, the LdapServer service will be restarted. After the LdapServer service is restarted, log in to MRS Manager again.

****End

**Summary and Suggestions**

It is recommended that software installed by the user be separated from system software. A system software upgrade may cause compatibility problems.
4.1 Failed to Call an API to Create a Cluster

Issue

After the user follows instructions in Creating a Cluster and Running a Job to call the cluster creation API POST /v1.1/{project_id}/run-job-flow, 500 is returned.

Symptom

After the cluster creation API POST /v1.1/{project_id}/run-job-flow is called, 500 is returned.

Cause Analysis

The user does not have permission to create a cluster.

Procedure

Check whether the user has permissions to create an MRS cluster. For details, see Permissions Management.
5 Cluster Management

5.1 Failed to Scale In Task Nodes

Issue

The customer resizes the cluster on the MRS cluster details page and changes the number of Task nodes to 0, with a result that Task nodes are not scaled in.

Symptom

When a customer adjusts the Task nodes on the MRS cluster details page, a message stating "This operation is not allowed because the number of instances of NodeManager will be less than the minimum configuration after scale-in, which may cause data loss." is displayed, indicating that cluster scale-in fails.

Cause Analysis

The customer stops the NodeManager service of a Core node. As a result, if all Tasks are found to be decommissioned when MRS checks whether Task nodes can be decommissioned, no NodeManager is available, and so is the Yarn service. However, MRS determines that Task nodes can be decommissioned only when the remaining NodeManagers must be greater than or equal to 1.

Procedure

Step 1 Log in to the MRS management console. On the Active Clusters page, click the cluster name. The cluster details page is displayed.

Step 2 Click Manage next to Cluster Manager to access the MRS Manager page.

Step 3 On the MRS Manager page, choose Services > Yarn > Instance.

Step 4 Select the NodeManager instance of the Core node and choose More > Start Instance.
Step 5  Then, scale in Task nodes on the cluster list page.

1. Click the cluster name. On the cluster details page that is displayed, click the Nodes tab.
2. Locate the row that contains the Task node group and click Scale In in the Operation column.

![Figure 5-1 Scale-in Task](image)

3. Click OK. In the displayed dialog box, click Yes.

Step 6  If you do not want to use NodeManager of the Core node after the Task node is successfully scaled in, stop NodeManager.

----End

Summary and Suggestions

Typically, NodeManager of the Core node is not stopped. You are not advised to change the cluster deployment architecture.

5.2 Expired OBS Certificate in a Cluster

Issue

The certificate expired when the customer accesses OBS from the MRS cluster.

Symptom

The certificate expiration problem occurs when the customer accesses the OBS service in the MRS cluster, resulting in failure to access the OBS system data.

Cause Analysis

The certificate generated by the OBS system has a validity period. After the validity period expires, the server automatically updates the certificate. As a result, an error occurs when the customer uses the old certificate to access the OBS system.
Procedure

Log into the cluster node in the background using VNC and run the following command. For details about the configuration of each Region on HUAWEI CLOUD, see Table 5-1.

```
/opt/Bigdata/jdk/bin/keytool -delete -storepass changeit -alias ${uds_url} -keystore /opt/Bigdata/jdk/jre/lib/security/cacerts || true
echo | /usr/bin/openssl s_client -connect ${uds_url}:${uds_port} 2>&1 | sed -ne '/-BEGIN CERTIFICATE-/,/-END CERTIFICATE-/p' > /tmp/obs.pem
/usr/bin/openssl x509 -in /tmp/obs.pem -text | grep CN
yes/opt/Bigdata/jdk/bin/keytool -import -storepass changeit -alias ${uds_url} -keystore /opt/Bigdata/jdk/jre/lib/security/cacerts -file /tmp/obs.pem
rm -rf /tmp/obs.pem
```

<table>
<thead>
<tr>
<th>Region</th>
<th>uds_url</th>
<th>uds_port</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN North-Beijing1</td>
<td>obs.cn-north-1.myhuaweicloud.com</td>
<td>443</td>
</tr>
<tr>
<td>CN North-Beijing4</td>
<td>obs.cn-north-4.myhuaweicloud.com</td>
<td>443</td>
</tr>
<tr>
<td>CN East-Shanghai1</td>
<td>obs.cn-east-3.myhuaweicloud.com</td>
<td>443</td>
</tr>
<tr>
<td>CN East-Shanghai2</td>
<td>obs.cn-east-2.myhuaweicloud.com</td>
<td>443</td>
</tr>
<tr>
<td>CN South-Guangzhou</td>
<td>obs.cn-south-1.myhuaweicloud.com</td>
<td>443</td>
</tr>
<tr>
<td>AP-Hong Kong</td>
<td>obs.ap-southeast-1.myhuaweicloud.com</td>
<td>443</td>
</tr>
<tr>
<td>AP-Bangkok</td>
<td>obs.ap-southeast-2.myhuaweicloud.com</td>
<td>443</td>
</tr>
<tr>
<td>AP-Singapore</td>
<td>obs.ap-southeast-3.myhuaweicloud.com</td>
<td>443</td>
</tr>
<tr>
<td>AF-Johannesburg</td>
<td>obs.af-south-1.myhuaweicloud.com</td>
<td>443</td>
</tr>
</tbody>
</table>
5.3 Adding a New Disk to an MRS Cluster

**Issue**

MRS HBase is unavailable.

**Symptom**

High disk usage of the user’s host causes service faults.

**Possible Cause**

Services are unavailable due to insufficient disk capacity.

**Procedure**

**Step 1** Purchase an EVS disk. For details, see Purchasing an EVS Disk.

**Step 2** Attach the EVS disk. For details, see Attaching a Non-Shared Disk.

- If the EVS disk has been attached, go to Step 6.
- If an ECS fails to be selected when you attach the EVS disk on the EVS management console, go to Step 3.

**Figure 5-2** Failed to select the cloud server.

**Step 3** Log in to the ECS management console and click the name of the ECS to which new disks are about to be attached.

**Step 4** On the Disks tab page, click Attach Disk.
**Step 5**  Select new disks to be attached and click **OK**.

**Step 6**  Initialize Linux data disks. For details, see *Initializing a Linux Data Disk (fdisk)*.
The attachment point directory is the existing DataNode instance ID plus one. For example, if you run the `df -h` command and find that the existing ID is `/srv/BigData/hadoop/data1`, the new attachment point is `/srv/BigData/hadoop/data2`. When initializing a Linux data disk to create an attachment point, name the new attachment point `/srv/BigData/hadoop/data2` and attach a new partition to the attachment point. Example:

```
mkdir /srv/BigData/hadoop/data2
mount /dev/xvdb1 /srv/BigData/hadoop/data2
```

**Step 7** Run the following command to grant the `omm` user permission to the new disk:
```
chown omm:wheel New attachment point
```
Example: `chown omm:wheel /srv/BigData/hadoop/data2`

**Step 8** Run the `chmod 701` command to grant the execution permission on the new attachment point directory.
```
chmod 701 New attachment point
```
Example: `chmod 701 /srv/BigData/hadoop/data2`

**Step 9** Log in to MRS Manager and add data disks to DataNode and NodeManager instances.

**Step 10** Choose Services > HDFS > Instance > DataNode > Instance Configuration. Set Type to All, and modify DataNode instance configurations of the current node.
- Enter `dfs.datanode.data.dir` in the search box and modify the parameter value to `/srv/BigData/hadoop/data1/dn,/srv/BigData/hadoop/data2/dn`.

If the values of the two parameters have been modified, click Save Configuration and select Restart role instance to restart the DataNode instance.

**Step 11** Choose Services > Yarn > Instance > NodeManager > Instance Configuration. Set Type to All, and modify the Yarn NodeManager instance configurations of the current node.
- Enter `yarn.nodemanager.local-dirs` in the search box and modify the parameter value to `/srv/BigData/hadoop/data1/nm/localdir,/srv/BigData/hadoop/data2/nm/localdir`.
- Enter `yarn.nodemanager.log-dirs` in the search box and modify the parameter value to `/srv/BigData/hadoop/data1/nm/containerlogs,/srv/BigData/hadoop/data2/nm/containerlogs`.

If the values of the two parameters have been modified, click Save Configuration and select Restart role instance to restart the NodeManager instance.

**Step 12** To check whether capacity expansion is successful, choose Services > HDFS > Instance > DataNode and check whether the total disk capacity in the DataNode Capacity real-time monitoring metric has been increased in the Charts area. If the DataNode Capacity monitoring metric does not exist in the Charts area, click Customize to add the metric.
- If the total disk capacity has been increased, the capacity expansion is complete.
• If the total disk capacity does not increase, contact Huawei technical support.

**Step 13** (Optional) Add data disks to a Kafka instance.

Modify the Kafka instance configurations of the current node.

1. Log in to MRS Manager, and choose **Services > Kafka > Instance > Broker > Instance Configuration**. Set **Type** to **All**.
2. Enter **log.dirs** in the search box to add a new disk and use commas (,) to separate it from the existing one.

   For example, if there is only one existing Kafka data disk and a new one is added, change `/srv/BigData/kafka/data1/kafka-logs` to `/srv/BigData/kafka/data1/kafka-logs,/srv/BigData/kafka/data2/kafka-logs`.

3. Click **Save Configuration** and select **Restart role instance** to restart the instance as prompted.
4. To check whether capacity expansion is successful, choose **Services > Kafka > Instance > Broker** and check whether the total disk capacity in the **Capacity of Broker Disks** real-time monitoring metric has been increased.

---End

--- CAUTION ---

After the disk capacity of a cluster node is expanded, if a new node is added to the cluster, you need to add disks to the new node by referring to the procedure. Otherwise, data may be lost.

--- Summary and Suggestions ---

• If the disk usage exceeds 85%, you are advised to expand disk capacity and attach the newly purchased disks to an ECS to associate with a cluster.
• Perform the attachment steps and set parameters based on the site requirements.

## 5.4 Replacing a Disk in an MRS Cluster

### Issue

A disk is not accessible.

### Symptom

A customer purchased an MRS cluster with local disks. A Core node in the cluster has disk faults, resulting in file read failures.

### Cause Analysis

The disk hardware is faulty.
Procedure

**NOTE**

This procedure is applicable to analysis clusters. If you need to replace disks for a streaming cluster or hybrid cluster, contact HUAWEI CLOUD technical support.

Follow this procedure when replacing disks for an analysis cluster, or contact technical support when replacing disks for a streaming cluster.

**Step 1** Log in to **MRS Manager**.

**Step 2** Choose **Hosts**, click the name of the host to be decommissioned, click **RegionServer** in the **Roles** list, and choose **More > Decommission**.

**Step 3** Choose **Hosts**, click the name of the host to be decommissioned, click **DataNode** in the **Roles** list, and choose **More > Decommission**.

**Step 4** Choose **Hosts**, click the name of the host to be decommissioned, click **NodeManager** in the **Roles** list, and choose **More > Decommission**.

**NOTE**

If other instances exist on this host, perform the following steps to decommission the instances.

**Step 5** Run the vim `/etc/fstab` command to comment out the mount point of the old disk.

**Figure 5-5** Commenting out the mount point of the old disk

```
[root@node-ana coregenx0001 -]# vim /etc/fstab

# devpts /dev/pts devpts mode=0220,gid=5 0 0
proc /proc proc defaults 0 0
sysfs /sys sysfs noatime 0 0
debugfs /sys/kernel/debug debugfs noauto 0 0
tmpfs /run tmpfs noatime 0 0
/dev/sdb1 /srv/BigData/hadoop/data1 ext4 defaults,noatime,nodiratime 1 0
```

**Step 6** Migrate the user data on the old disk (for example, `/srv/BigData/hadoop/data1/`).

**Step 7** On the cluster details page, click the **Nodes** tab.

**Step 8** Click the node where the disk is to be replaced to go to the ECS management console. Click **Stop** to shut down the node.

**Step 9** Contact HUAWEI CLOUD technical support to replace the disk in the background.

**Step 10** On the ECS management console, click **Start** to start the node where the disk has been replaced.

**Step 11** Run the `fdisk -l` command to view the new disk.

**Step 12** Run the `cat /etc/fstab` command to obtain the drive letter.
Step 13 Use the corresponding drive letter to format the new disk.
Example: mkfs.ext4 /dev/sdh

Step 14 Run the following command to attach the new disk.
mount new disk mount point
Example: mount /dev/sdh /srv/Bigdata/hadoop/data1

Step 15 Run the following command to grant the omm user the permissions to access the new disk:
chown omm:wheel mount point
Example: chown -R omm:wheel /srv/Bigdata/hadoop/data1

Step 16 Add the UUID of the new disk to the fstab file.
1. Run the blkid command to check the UUID of the new disk.

```
# blkid \
/mnt/new_disk
```

2. Open the /etc/fstab file and add the following information:
```
UUID=New disk UUID /srv/BigData/hadoop/data1 ext4 defaults,noatime,nodiratime 1 0
```

Step 17 Log in to MRS Manager.

Step 18 Choose Hosts, click the name of the host to be decommissioned, click RegionServer in the Roles list, and choose More > Recommission.

Step 19 Choose Hosts, click the name of the host to be decommissioned, click DataNode in the Roles list, and choose More > Recommission.

Step 20 Choose Hosts, click the name of the host to be decommissioned, click NodeManager in the Roles list, and choose More > Recommission.

**NOTE**
If other instances exist on this host, perform the following steps to recommission the instances.

Step 21 Choose Services > HDFS. In the HDFS Summary area on the Service Status tab page, check whether the number of Missing Blocks is 0.
- If the number is 0, no further action is required.
- If the number is not 0, contact HUAWEI CLOUD technical support.

----End
5.5 MRS Backup Failure

Issue
MRS backup keeps failing.

Symptom
MRS backup keeps failing.

Possible Cause
The backup directory is connected to the system disk using a soft link. As a result, if the system disk is full, the backup fails.

Procedure
Step 1 Check whether the backup directory is connected to the system disk using a soft link.
1. Log in to the active and standby Master nodes in the cluster as user root.
2. Run the df -h command to check the storage usage of the system disk.
3. Run the ll /srv/BigData/LocalBackup command to check whether the backup directory is connected to /opt/Bigdata/LocalBackup using a soft link.
   Check whether the backup file is connected to the system disk using a soft link and whether the system disk has sufficient space. If the soft link is used for connecting to the system disk and the system disk space is insufficient, go to Step 2. If the soft link is not used, the failure is not caused by insufficient system disk space. Contact technical support for troubleshooting.

Step 2 Move historical backup data to a new directory on the data disk.
1. Log in to the Master node as user root.
2. Run the su - omm command to switch to user omm.
3. Run the rm -rf /srv/BigData/LocalBackup command to delete the soft link of the backup directory.
4. Run the mkdir -p /srv/BigData/LocalBackup command to create a backup directory.
5. Run the mv /opt/Bigdata/LocalBackup/* /srv/BigData/LocalBackup/ command to move the historical backup data to the new directory.
5.6 Inconsistency Between df and du Command Output on the Core Node

**Issue**

The capacity displayed in the df command output on the Core node is inconsistent with that displayed in the du command output.

**Symptom**

The capacity displayed in the df command output on the Core node is different from that displayed in the du command output.

The disk usage of the /srv/BigData/hadoop/data1/ directory queried by running the df -h command differs greatly from that queried by running the du -sh /srv/BigData/hadoop/data1/ command. The difference is greater than 10 GB.

**Cause Analysis**

The lsol |grep deleted command output indicates that a large number of log files in the directory are in the deleted state.

When some Spark tasks are running for a long time, some containers in the tasks keep running and logs are continuously generated. When printing logs, the executor of Spark uses the log4j log scrolling function to output logs to the stdout file. The container also monitors this file. As a result, the file is monitored by two processes at the same time. When one process scrolls according to the configuration, the earliest log file is deleted, but the other process still occupies the file handle. As a result, a file in the deleted state is generated.

**Procedure**

Change the output directory name for executor logs of Spark.

1. Open the log configuration file. By default, the configuration file is located in <Client address>/Spark/spark/conf/log4j-executor.properties.
2. Change the name of the log output file. For example, change log4j.appender.sparklog.File = $spark.yarn.app.container.log.dir$/stdout to log4j.appender.sparklog.File = $spark.yarn.app.container.log.dir$/stdout.log.
3. Save the configuration and exit.
4. Submit the tasks again.

5.7 Disassociating a Subnet from the ACL Network

**Scenario**

You can disassociate a subnet from the ACL network when necessary.
Procedure

Step 1  Log in to the management console.

Step 2  On the console homepage, under Network, click Virtual Private Cloud.

Step 3  In the navigation tree on the left, choose Network ACL.

Step 4  Locate the target network ACL in the right pane, and click the network ACL name to switch to the network ACL details page.

Step 5  On the displayed page, click the Associated Subnets tab.

Step 6  On the Associated Subnets page, locate the target network ACL and click Disassociate in the Operation column.

Step 7  Click OK.

----End

5.8 MRS Becomes Abnormal After hostname Modification

Issue

What should I do if MRS becomes abnormal after hostname is modified?

Symptom

MRS becomes abnormal after hostname is modified.

Possible Cause

The hostname modification causes compatibility problems and faults.

Procedure

Step 1  Log in to any node in the cluster as user root.

Step 2  Run the cat /etc/hosts command on the node to check the value of hostname of each node and set the newhostname variable based on the value.

Step 3  Run the sudo hostnamectl set-hostname ${newhostname} command on the node where hostname is modified to restore the correct host name.

\[\text{NOTE}\]

\[\text{${newhostname}}\]: new value of hostname

Step 4  After the modification, log in to the node where hostname is modified again and check whether the new host name takes effect.

----End
5.9 Determining Whether a Process Is Killed

Symptom

A DataNode is restarted unexpectedly. No operation is performed to manually restart the DataNode. You need to identify the process that kills the DataNode service.

Cause Analysis

Common reasons for abnormal process termination are as follows:

- **OOM of the Java process is killed.**
  In general, the OMM Killer is configured for Java processes to detect and kill OOM. The OOM log is printed in the out log. In this case, you can view the run log (for example, the DataNode's log path is /var/log/Bigdata/hdfs/dn/hadoop-omm-datanode-host name.log) to check whether OutOfMemory is printed.

- **DataNode is killed by other processes or manually.**
  Check the DataNode run log file (/var/log/Bigdata/hdfs/dn/hadoop-omm-datanode-host name.log). It is found that the health check fails after "RECEIVED SIGNAL 15" is received. In the following example, the DataNode is killed at 11:04:48 and then started at 11:06:52 two minutes later.

  2018-12-06 11:04:48,433 | ERROR | SIGTERM handler | RECEIVED SIGNAL 15: SIGTERM |
  LogAdapter.java:69
  2018-12-06 11:04:48,436 | INFO  | Thread-1 | SHUTDOWN_MSG: /************************************************************
  ************************************************************/ | LogAdapter.java:45
  2018-12-06 11:06:52,744 | INFO  | main | STARTUP_MSG:

  According to the logs, DataNode was closed and then the health check reported the exception. After 2 minutes, NodeAgent started the DataNode process.

Procedure

Add the rule for recording the kill command in the audit log of the operating system. The process that delivers the kill command will be recorded in the audit log.

Operation impact

- Printing audit logs affects operating system performance. However, analysis result shows that the impact is less than 1%.
- Printing audit log occupies some disk space. The logs to be printed are within megabytes. By default, the aging mechanism and the mechanism for checking the remaining disk space are configured. Therefore, the disk space will not be used up.

Locating Method

Perform the following operations on nodes that may restart the DataNode process:
**Step 1** Log in to the node as the root user and run the service auditd status command to check the service status.

Checking for service auditd running

If the service is not started, run the service auditd restart command to restart the service. The command execution takes less than 1 second and has no impact on the system.

Shutting down auditd done
Starting auditd done

**Step 2** The audit rule of the kill command is temporarily added to audit logs.

Add an audit rule:

```bash
auditctl -a exit,always -F arch=b64 -S kill -S tkill -S tgkill -F a1!=0 -k process_killed
```

View the rule:

```bash
auditctl -l
```

**Step 3** If a process is killed due to an exception, you can run the ausearch -k process_killed command to query the kill history.

```bash

```

**NOTE**

Note: a0 is the PID (hexadecimal) of the process that is killed, and a1 is the semaphore of the kill command.

----End

**Verification**

**Step 1** Restart an instance of the node on MRS Manager, for example, DataNode.

**Step 2** Run the ausearch -k process_killed command to check whether logs are printed.

The following is an example of the ausearch -k process_killed |grep ".sh" command. The command output indicates that the hdfs-daemon-ada* script closed the DataNode process.

```

```

----End

Stop auditing the kill command.

**Step 1** Run the service auditd restart command. The temporarily added kill command audit logs are cleared automatically.
Step 2 Run the `auditctl -l` command. If no information about killing a process is returned, the rule is cleared successfully.

---End

5.10 Failed to Configure Cross-Cluster Mutual Trust

Symptom
The cross-cluster mutual trust relationship cannot be established between a cluster earlier than MRS 1.8.2 and a cluster later than MRS 1.8.2.

Cause Analysis
After cross-cluster mutual trust relationship is configured, internal users `krbtgt/local cluster domain name@external cluster domain name` and `krbtgt/external cluster domain name@local cluster domain name` are added to the two clusters. In versions earlier than MRS 1.8.2, the default user password is `Admin@123`, in MRS 1.8.2 or later, the default user password is `Crossrealm@123`. The cross-cluster mutual trust configuration fails because the user passwords of the two clusters are different.

Procedure
- **Scenario without mutual trust being configured:**
  a. Before configuring the mutual trust, log in to the Master node in the cluster of MRS 1.8.2 or later.
  b. Change the value of `local cross_realm_account_pwd=${DEFAULT_CROSS_REALM_PWD}` in the `add_cross_realm_princ` method of the `/opt/Bigdata/om-0.0.1/sbin/addRealm.sh` script on all Master nodes to `local cross_realm_account_pwd=${DEFAULT_PWD}` (in line 1001 of the script).

  **NOTE**

  Perform steps from a to b on all Master nodes in the cluster of MRS 1.8.2 or later.
  c. Then, configure cross-cluster mutual trust by referring to Configuring Cross-Cluster Mutual Trust Relationships
  d. Check whether the mutual trust relationship is established.
     - If yes, the configuration is complete.
     - If the relationship fails to be established, refresh the client configuration and check whether the trust relationship is established. If the problem persists, submit a service ticket.

- **Scenario with mutual trust being configured**
  a. Log in to the Master node in the cluster of MRS 1.8.2 or later.
  b. Run the `/home/omm/kerberos/bin/kadmin -p kadmin/admin` command and enter the password of the Kerberos client.
c. Run the `listprincs` command and press Enter to query user information.

d. Run the `delprinc` command to delete users `krbtgt/local cluster domain name`@`external cluster domain name` and `krbtgt/external cluster domain name`@`local cluster domain name`.

e. Run the `quit` command to exit the Kerberos client.

f. Change the value of `local cross_realm_account_pwd="$ DEFAULT_CROSS_REALM_PWD"` in the `add_cross_realm_princ` method of the `/opt/Bigdata/om-0.0.1/sbin/addRealm.sh` script on the Master nodes to `local cross_realm_account_pwd="$ DEFAULT _PWD"` (in line 1001 of the script).

g. Log in to MRS Manager, and choose Services.

h. Choose More > Synchronize Configuration.

i. In the displayed dialog box, select Restart the service or instance whose configuration has expired. and click OK to restart the service whose configuration has expired.

   During configuration synchronization, the `addRealm.sh` script is invoked to add the `krbtgt` user.

   □ □ NOTE

   Perform steps from a to i on all Master nodes in the cluster of MRS 1.8.2 or later.

j. Check whether the mutual trust is established. If it still fails, submit a service ticket.

5.11 Network Is Unreachable When Using `pip3` to Install the Python Package in an MRS Cluster

Issue

When the Python package is installed using `pip3`, an error message is displayed, indicating that the network is unreachable.
Symptom
When a user runs the pip3 install command to install the Python package, an error message is displayed, indicating that the network is unreachable. For details, see the following figure:

![Error Message]

Cause Analysis
The customer does not bind an EIP to the Master node.

Procedure

**Step 1** Log in to the MRS management console.

**Step 2** Choose Clusters > Active Clusters, select the faulty cluster, and click its name to check the Basic Information on the Dashboard tab page.

**Step 3** On the Nodes tab page, click the name of a Master node in the Master node group to log in to the ECS management console.

**Step 4** Click the EIPs tab and click Bind EIP to bind an EIP to the ECS.

**Step 5** Log in to the Master node and run the pip3 install command to install the Python package.

5.12 Connecting the Open-Source confluent-kafka-go to the Security Cluster of MRS

**Issue**
How do I connect the Open-Source confluent-kafka-go an MRS security cluster?

**Symptom**
The open-source confluent-kafka-go fails to connect to the security cluster of MRS.

**Cause Analysis**
By default, the librdkafka library on which confluent-kafka-go depends uses the host name of the broker as a part of the server principle. As a result, the authentication fails.

**Procedure**
The procedure for modifying the librdkafka is as follows:
1. The librdkafka source code address: https://github.com/edenhill/librdkafka

2. Add the `sasl.kerberos.service.name` configuration item to the src/rdkafka_conf.c file.

```c
"Kerberos principal name that Kafka runs as.", _sdef = "kafka" ),
"sasl.kerberos.principal", _RK_C_STR, _RK(sasl.principal), "This client’s Kerberos principal name.", _sdef = "kafkaclient" ), + { _RK_GLOBAL, _RK_C_STR, _RK(sasl.domain_name), + "" " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " 
```

3. Add the `domain_name` field to the src/rdkafka_conf.h file.

```c
--- src/rdkafka_conf.h  2017-10-17 11:20:56.000000000 +0800 +++ src/rdkafka_conf.h 2017-10-25 16:26:34.000000000 +0800 @@ -118,12 +118,13 @@
 struct {
   const struct rd_kafka_sasl_provider *provider;
   char *principal;
   char *mechanisms; +
   char *domain_name;
   char *service_name;
   char *kinit_cmd;
   char *keytab;
   int   relogin_min_time;
   char *username;
   char *password; #if WITH_SASL_SCRAM
```

4. Replace `hostname` with `domainName` in the src/rdkafka_sasl_cyrus.c file.

```c
 rk->conf.sasl.mechanisms,
 rk->conf.api_version_request ?
 "": try api.version_request=true();
 -rd_strdupa(&hostname, rktrans->rktrans_rkb->rkb_nodename); +
 //rd_strdupa(&hostname, rktrans->rktrans_rkb->rkb_nodename); +
 rd_strdupa(&hostname, rk->conf.sasl.domain_name); +
 if ((t = strchr(hostname, ´:´))-remove "" /* remove "":port */
```

5. Recompile librdkafka (ensure that libsasl2-dev has been installed). For details, see https://github.com/edenhill/librdkafka/tree/v0.11.1.

```
./configure
make
make install
```

6. Add the following configuration items when using the client:

```c
"security.protocol": "SASL_PLAINTEXT",
"sasl.kerberos.service.name": "kafka",
"sasl.kerberos.keytab":="/opt/nemon/user.keytab",
"sasl.kerberos.principal": "nemon@HADOOP.COM",
"sasl.kerberos.domain.name": "hadoop.hadoop.com",
```

**NOTE**: This is not automatically used but must be added to the "template in sasl.kerberos.kinit.cmd as "

```
NOTE
- sasl.kerberos.keytab: On MRS Manager, choose System > Manage User. Locate the row that contains the target user, choose More > Download authentication credential. Save the file and decompress it to extract the user.keytab file.
- sasl.kerberos.principal: Enter the actual user name.
- sasl.kerberos.domain.name: The domain naming rule is hadoop. toLowerCase(realm): If the cluster domain name (default_realm) is huawei.com, the value of domain is hadoop.huawei.com. On MRS Manager, choose Services > KrbServer > Service Configuration > All, and search for and view the value of default_realm.
```

---
5.13 Failed to Periodically Back Up an MRS 1.7.2 Cluster

Issue
An MRS cluster fails to be backed up periodically.

Symptom
An MRS 1.7.2 cluster fails to be backed up periodically, and alarm ALM-12034 Periodic Backup Failure is displayed on the Alarms page of MRS Manager.

Cause Analysis
During periodic backup of the cluster, a soft link is generated to connect /srv/BigData/LocalBackup to /opt/Bigdata/LocalBackup, which uses the disk space of the root directory. Because the disk space of the root directory is insufficient, backup files cannot be written into the root directory. As a result, the periodic backup of the cluster fails.

Procedure
1. Log in to the active and standby Master nodes.
2. Run the cd /srv/BigData/ command to go to the directory where the backup file is stored.
3. Run the unlink LocalBackup command to delete the LocalBackup soft link.
4. Run the mkdir –p LocalBackup command to create the LocalBackup directory.
5. Run the chown –R omm:wheel LocalBackup command to change the user and group to which the file belongs.
6. Run the chmod 700 LocalBackup command to modify the read and write permissions on the file.
7. Log in to MRS Manager and perform periodic backup again.

5.14 Failed to Download the MRS Cluster Client

Issue
On the local Master host, a user attempts to download an MRS cluster client for another remote host. However, the system displays a message indicating that the network or parameter is abnormal.

Symptom
On the local Master host, a user attempts to download an MRS cluster client for another remote host. However, the system displays a message indicating that the network or parameter is abnormal.
Cause Analysis

- The two hosts are in different VPCs.
- The password is incorrect.
- The firewall is enabled on the remote host.

Procedure

- The two hosts are in different VPCs.
  Enable port 22 of the remote host.
- The password is incorrect.
  Check whether the password is correct. The password cannot contain special characters.
- The firewall is enabled on the remote host.
  Download the MRS cluster client to the server and run the `scp` command provided by Linux to remotely send the client to the remote host.

5.15 An Error Is Reported When a Flink Job Is Submitted in a Cluster with Kerberos Authentication Enabled

Issue

An error is reported when a user submits a Flink job with Kerberos authentication enabled.

Symptom

When a user submits the official Flink case `./flink run /opt/client/Flink.flink/examples/streaming/WordCount.jar`, the error message “unable to establish the security context” is displayed.

Cause Analysis

1. Kerberos authentication is enabled but jobs cannot be submitted. Therefore, the permission configuration is checked first. It is found that the parameters in `/Flink/flink/conf/flink-conf.yaml` are not correctly configured.
Figure 5-7 flink-conf.yaml configuration

For details, see Configuring and Managing Flink.

2. After the configuration is modified and refreshed, the job can be submitted again, but the "log4j:ERROR setFile(null,ture) call failed" error is reported.

Figure 5-8 log4j error

3. Check on log4j indicates that the name of the log4j.properties file is changed to log4g-cli.properties (the name of log4j.properties is fixed and cannot be changed). Therefore, an error is reported.

Figure 5-9 Viewing the log4j file

4. After the modification, the job can be submitted properly.

Figure 5-10 Job submission

Procedure

Step 1  Determine the place to submit jobs using the client, outside the cluster or in the cluster.
1. If you want to submit jobs using the client in the cluster, switch to user omm before submitting jobs.
2. If you want to submit jobs using the client outside the cluster, switch to user root before submitting jobs.

Step 2 Check whether the parameters in the flink-conf.yaml file are correctly configured. For details, see Configuring and Managing Flink.

Step 3 For clusters with Kerberos authentication enabled, the configuration items include keytab and principal of Kerberos.
   - Download the user keytab file from the KDC server, and place the keytab file to a folder on the host of the Flink client (for example, /home/flinkuser/keytab).
   - Configure the following parameters in the ${FLINK_HOME}/conf/flink-conf.yaml file:
     a. Keytab path (note that there is a space before the parameter):
        security.kerberos.login.keytab: /home/flinkuser/keytab/uer.keytab
     b. Principal name (developer username):
        security.kerberos.login.principal:flinkuser

Step 4 Submit the job ./flink run /opt/client/Flink/flink/examples/streaming/WordCount.jar again and check whether the job can be submitted.
   - If the job can be submitted, the permission authentication is correct. In this case, you can check other errors. In this example, the name of log4j.properties is changed. After the name is restored, the job can be submitted normally.
   - If the job fails to be submitted, submit a service ticket.

----End

Related Information
For details about how to use Flink, see Using Flink from Scratch.

5.16 Scale-out Failure

Issue
MRS cluster scale-out fails when the console page is normal.

Symptom
The MRS console is normal, and no alarm or error message is displayed on MRS Manager. However, an error message is displayed during cluster scale-out, indicating that the MRS cluster contains nodes that are not running and asking you to try again later.

Cause Analysis
MRS cluster scale-in and scale-out can be performed only when the cluster is running properly. Therefore, you need to check whether the cluster is normal.
Currently, a message is displayed indicating that there are nodes that are not running in the cluster. However, the console and MRS Manager pages are normal. Therefore, the possible cause is that the cluster status in the database is abnormal or is not updated. As a result, the nodes in the cluster are not in the normal state, causing the failure.

**Procedure**

**Step 1** Log in to the MRS management console and click the cluster name to go to the cluster details page. Check the cluster status and ensure that the cluster is in the **Running** state.

**Step 2** Click **Nodes** to view the status of all nodes. Ensure that all nodes are in the **Running** state.

**Step 3** Log in to the podMaster node in the cluster, switch to the deployer node of MRS, and view the **api-gateway.log** file.

1. Run the `kubectl get pod -n mrs` command to view the **pod** of the deployer node corresponding to MRS.
2. Run the `kubectl exec -ti $({pod of the deployer node}) -n mrs /bin/bash` command to log in to the corresponding pod. For example, run the `kubectl exec -ti mrsdeployer-78bc8c76cf-mn9ss -n mrs /bin/bash` command to access the deployer container of MRS.
3. In the `/opt/cloud/logs/apigateway` directory, view the latest `api-gateway.log` file and search for key information (such as *ERROR*, *scaling*, *clusterScaling*, *HostState*, *state-check*, or cluster ID) in the file to check the error type.
4. Rectify the fault based on the error information and perform the scale-out again.
   - If the scale-out is successful, no further action is required.
   - If the scale-out fails, go to **Step 4**.

**Step 4** Run the `/opt/cloud/mysql -u$Username -P$Port -h$Address -p$Password` command to log in to the database.

**Step 5** Run the `select cluster_state from cluster_detail where cluster_id="Cluster ID"` command to check the value of **cluster_state**.

- If the value of **cluster_state** is 2, the cluster status is normal. Go to **Step 6**.
- If the value of **cluster_state** is not 2, the cluster status in the database is abnormal. You can run the `update cluster_detail set cluster_state=2 where cluster_id="Cluster ID"` command to refresh the cluster status and check the value of **cluster_state**.
  - If the value of **cluster_state** is 2, the cluster status is normal. Go to **Step 6**.
  - If the value of **cluster_state** is not 2, submit a service ticket.

**Step 6** Run the `select host_status from host where cluster_di="Cluster ID"` command to query the cluster host status.

- If the host is in the started state, no further action is required.
- If the host is not in the started state, run the `update host set host_status='started' where cluster_id="Cluster ID"` command to update the host status to the database.
– If the host is in the started state, no further action is required.
– If the host is not in the started state, submit a service ticket.

5.17 Error Occurs When MRS Executes the Insert Command Using Beeline

Issue
An error occurs when MRS executes the insert command using Beeline.

Symptom
When the `insert into` statement is executed in Beeline of Hive, the following error is reported:

```
Mapping run in Tez on Hive transactional table fails when data volume is high with error:
"org.apache.hadoop.hive.ql.lockmgr.LockException Reason: Transaction... already aborted, Hive SQL state [42000]."
```

Cause Analysis
This problem is caused by improper cluster configuration and Tez resource setting.

Procedure
This problem can be solved by setting configuration parameters on Beeline.

Step 1  Set the following properties to optimize performance (you are advised to change them at the cluster level):
- Set `hive.auto.convert.sortmerge.join` to `true`.
- Set `hive.optimize.bucketmapjoin` to `true`.
- Set `hive.optimize.bucketmapjoin.sortedmerge` to `true`.

Step 2  Modify the following content to adjust the resources of Tez:
- Set `hive.tez.container.size` to the size of the Yarn container.
- Set `hive.tez.container.size` to the Yarn container size `yarn.scheduler.minimum-allocation-mb` or a smaller value (for example, a half or quarter of the Yarn container size). Ensure that the value does not exceed `yarn.scheduler.maximum-allocation-mb`.

---- End
6 Using HBase

6.1 Slow Response to HBase Connection

Issue

Under the same VPC network, response is slow when an external cluster connects to HBase through Phoenix.

Symptom

Under the same VPC network, response is slow when an external cluster connects to HBase through Phoenix.

Possible Cause

DNS has been configured. When a client connects to HBase, DNS resolves the server first, causing slow response.

Procedure

**Step 1**  Log in to the Master node as user root.

**Step 2**  Run the `vi /etc/resolv.conf` command to open the `resolv.conf` file and comment out the address of the DNS server, for example, #1.1.1.1.

-----End
6.2 Failed to Authenticate the HBase User

Issue

Failed to authenticate the HBase user.

Symptom

Failed to authenticate the HBase user on the client. The following error information is displayed.

```
2019-05-13 10:53:09,975 ERROR [localhost-startStop-1] huaweiConfig.LoginUtil: perhaps cause 1 is (wrong password) keytab file and user not match, you can kinit -k -t keytab user in client server to check.
2019-05-13 10:53:09,975 ERROR [localhost-startStop-1] huaweiConfig.LoginUtil: perhaps cause 2 is (clock skew) time of local server and remote server not match, please check ntp to remote server.
```

Possible Cause

The version of the JAR file in the JDK used by the user is different from that of the JAR file authenticated by MRS.

Procedure

1. Log in to the Master1 node as user root.
2. Run the following command to check the JAR file authenticated by MRS:
   - `ll /opt/share/local_policy/local_policy.jar`
   - `ll /opt/Bigdata/jdk{version}/jre/lib/security/local_policy.jar`
3. Download the JAR package queried in step 2 to the local host.
4. Copy the downloaded JAR package to the local JDK directory `/opt/Bigdata/jdk/jre/lib/security`.
5. Run the `cd /opt/client/HBase/hbase/bin` command to go to the `bin` directory of HBase.
6. Run the `sh start-hbase.sh` command to restart HBase.

----End

6.3 RegionServer Start Failure Due to Port Occupation

Symptom

RegionServer is in the Concerning state on the MRS Manager page.
Cause Analysis

1. View the RegionServer log (/var/log/Bigdata/hbase/rs/hbase-omm-xxx.log).
2. Run the `lsof -i:21302` command (the port number of MRS 1.7.X and later versions is 16020) to view the PID. Based on the PID, check the process. It is found that the RegionServer port is occupied by DFSZkFailoverController.
3. The value of `/proc/sys/net/ipv4/ip_local_port_range` is 9000 65500. The temporary port range and the MRS port range overlap. This is because the preinstall operation is not performed during installation.

Solution

**Step 1** Run the `kill -9 DFSZkFailoverController pid` command to ensure that another port is bound with after a restart and restart the RegionServer in the Concerning state.

----End

6.4 HBase Startup Failure Due to Insufficient Memory on the Node

Symptom

The RegionServer service of HBase is always in the Concerning state.

Cause Analysis

1. Check the RegionServer log (/var/log/Bigdata/hbase/rs/hbase-omm-XXX.out). It is found that the following information is printed:
   ```
   There is insufficient memory for the Java Runtime Environment to continue.
   ```
2. Run the `free` command to check the memory. It is found that the available memory of the node is insufficient.

Solution

**Step 1** Locate why the memory is insufficient. It is found that some processes occupy too much memory or the server does not have sufficient memory.

----End

6.5 HBase Service Unavailable Due to Poor HDFS Performance

Symptom

The HBase component intermittently reports alarms indicating that the service is unavailable.
Cause Analysis

HDFS performance is low, causing health check timeout and the alarm is generated accordingly. You can perform the following operations:

1. Check the HMaster log (/var/log/Bigdata/hbase/hm/hbase-omm-xxx.log) and check whether system pause, jvm, and other GC related information is frequently displayed in the HMaster log.

2. You can determine whether the fault is caused by poor HDFS performance using the following methods:
   a. Use the client for verification. Run hbase shell command to log in to HBase command line, run the `list` command to verify the running duration.
   b. Enable the debug log function of the HDFS, and run the `hadoop fs –ls /XXX/XXX` command to check the paths of lower-layer directories to check the running duration.
   c. Print HMaster process jstack:
      ```
      su - omm
      jps
      jstack pid
      ```

3. Process jstack is always in the `DFSClient.listPaths` state.

Solution

**Step 1** If this alarm is caused by poor HDFS performance, check whether Impala in the earlier version causes poor performance or JournalNode was incorrectly deployed during initial deployment (more than three JournalNode nodes are deployed).

-----End
6.6 HBase Startup Failure Due to Inappropriate Parameter Settings

Symptom

After some parameters are modified, HBase cannot be started.

Cause Analysis

1. Check the HMaster log (/var/log/Bigdata/hbase/hm/hbase-omm-xxx.log). It is found that the total of hbase.regionserver.global.memstore.size and hfile.block.cache.size is greater than 0.8, which causes the startup failure. Therefore, adjust the parameter values to make sure that the total is less than 0.8.

2. Check the HMaster and RegionServer out logs (/var/log/Bigdata/hbase/hm/hbase-omm-xxx.out/var/log/Bigdata/hbase/rs/hbase-omm-xxx.out). It is found that Unrecognized VM option is displayed.

   Unrecognized VM option
   Error: Could not create the Java Virtual Machine. Error: A fatal exception has occurred. Program will exit.

   Check the GC_OPTS parameters. It is found that the parameters contain unnecessary spaces, for example, -D sun.rmi.dgc.server.gcInterval=0x7FFFFFFFFFFFFFE.

Solution

Step 1 After the MemStore and cache parameters are modified, the HBase service is restarted successfully.

Step 2 After the GC_OPTS parameters are modified, the HBase service is restarted successfully.

----End

6.7 RegionServer Start Failure Due to Residual Processes

Symptom

The HBase service fails to start, and an error is reported during the health check.

Cause Analysis

Check detailed information about HBase startup on the MRS Manager page. It is found that the previous process is not quit is displayed.
Solution

Step 1 Log in to the node and run the `ps -ef | grep HRegionServer` command in the background. A residual process is found.

Step 2 After confirming that the process can be killed, kill the process. If the process cannot be stopped by running the `kill` command, run the `kill -9` command to forcibly stop the process.

Step 3 Restart the HBase service.

----End

6.8 HBase Startup Failure Caused by Quota Settings on HDFS

Symptom

HBase fails to start.

Cause Analysis

Check the HMaster log (`/var/log/Bigdata/hbase/hm/hbase-omm-xxx.log`). It is found that "The DiskSpace quota of /hbase is exceeded" is displayed.

Solution

Step 1 Run the `df -h` command to check data directory space. It is found that the directory space is full. Delete unnecessary data to free up space.

Step 2 Expand the node to ensure that the data directory space is sufficient.

----End
6.9 HBase Startup Failure Caused by Initial Timeout Reservation

Symptom

In MRS 1.5.x versions, HBase fails to start and an initialization timeout message is displayed.

Cause Analysis

A large number of regions in transition exist but the default value of `hbase.master.initializationmonitor.timeout` is 300 seconds, during which initialization cannot be completed. As a result, the HMaster process exits.

Solution

**Step 1** Log in to MRS Manager and choose Services > HBase > Service Configuration. The HBase configuration page is displayed.

**Step 2** Set Type to All, search for the `hbase.master.initializationmonitor.timeout` parameter, and increase the parameter value (for example, change the value to 30 minutes).

**Step 3** Click Save Configuration and select Restart the affected services or instances. Click OK to restart the services.

---End

6.10 HBase Startup Failure Caused by Corrupted Version Files

Background

HBase fails to start.

Cause Analysis

1. The `hbase.version` file is read during HBase startup. However, the log indicates that a reading exception occurs.
2. The file cannot be viewed by running the `hadoop fs -cat /hbase/hbase.version` command. The file is corrupted.

Solution

**Step 1** Run the `hbase hbck -fixVersionFile` command to restore the file.

**Step 2** If the problem persists after performing **Step 1**, obtain the `hbase.version` file from another cluster of the same version and upload the file to replace the original one.

**Step 3** Restart the HBase service.

----End

6.11 High CPU Usage of RegionServer Even No Services Are Running

Background

RegionServer occupies a large number of CPU resources when no services are running.

Cause Analysis

1. Run the `top` command to obtain the CPU usage of RegionServer processes and check the IDs of processes with high CPU usage.

2. Obtain the CPU usage of threads under these processes based on the RegionServer process IDs.

   Run the `top -H -p <PID>` (replace it with the actual RegionServer process ID).

   As shown in the following figure, the CPU usage of some threads reaches 80%.

<table>
<thead>
<tr>
<th>USER</th>
<th>PR</th>
<th>NI</th>
<th>VIRT</th>
<th>RES</th>
<th>SHR</th>
<th>S</th>
<th>%CPU</th>
<th>%MEM</th>
<th>TIME+</th>
<th>COMMAND</th>
</tr>
</thead>
<tbody>
<tr>
<td>30623</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 90.4</td>
<td>9.2</td>
<td>811:27.39</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30630</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 90.4</td>
<td>9.2</td>
<td>775:23.29</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30641</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 88.6</td>
<td>9.2</td>
<td>774:01.90</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30624</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 86.8</td>
<td>9.2</td>
<td>751:11.96</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30631</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 86.8</td>
<td>9.2</td>
<td>858:11.06</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30634</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 86.8</td>
<td>9.2</td>
<td>815:40.30</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30648</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 86.8</td>
<td>9.2</td>
<td>984:19.21</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30646</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 81.5</td>
<td>9.2</td>
<td>731:40.75</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30622</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 79.7</td>
<td>9.2</td>
<td>739:26.82</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30637</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 79.7</td>
<td>9.2</td>
<td>784:40.04</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30627</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 79.0</td>
<td>9.2</td>
<td>852:51.22</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30638</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 76.2</td>
<td>9.2</td>
<td>920:20.73</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30629</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 74.4</td>
<td>9.2</td>
<td>1024:22</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30647</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 72.7</td>
<td>9.2</td>
<td>835:17.18</td>
<td>java</td>
<td></td>
</tr>
<tr>
<td>30644</td>
<td></td>
<td></td>
<td>34.2g</td>
<td>23g</td>
<td>23m</td>
<td>R 63.8</td>
<td>9.2</td>
<td>958:48.30</td>
<td>java</td>
<td></td>
</tr>
</tbody>
</table>

3. Obtain the thread stack information based on the ID of the RegionServer process.

   `jstack 12345 >allstack.txt` (Replace it with the actual RegionServer process ID.)
4. Convert the thread ID into the hexadecimal format:

```
printf "%x\n" 30648
```

In the command output, the TID is `77b8`.

5. Search the thread stack based on the hexadecimal TID. It is found that the compaction operation is performed.

6. Perform the same operations on other threads. It is found that the threads are compaction threads.

The threads that consume a large number of CPU resources are compaction threads. Some threads invoke the Snappy compression algorithm, and some threads invoke HDFS data writing and reading. Each region has massive sets of data and numerous data files and uses the Snappy compression algorithm. For this reason, the compaction operations consume a large number of CPU resources.

Solution

This is a normal phenomenon.

Fault Locating Methods

**Step 1** Run the `top` command to check the process with high CPU usage.

**Step 2** Check the threads with high CPU usage in the process.

Run the `top -H -p <PID>` command to print CPU usage of threads under the process.

Obtain the thread with the highest CPU usage from the query result. You can also obtain the thread by running the following command:

```
ps -mp <PID> -o THREAD,tid,time | sort -rn
```

View the command output to obtain the ID of the thread with the highest CPU usage.

**Step 3** Obtain the stack of the faulty thread.

The jstack tool is the most effective and reliable tool for locating Java problems. You can obtain the jstack tool from the `java/bin` directory.

```
jstack <PID> > allstack.txt
```
Obtain the process stack and output it to a local file.

**Step 4** Convert the thread ID into the hexadecimal format:

```bash
printf "%x\n" <PID>
```

The process ID in the command output is the TID.

**Step 5** Run the following command to obtain the TID and output it to a local file:

```bash
jstack <PID> | grep <TID> > Onestack.txt
```

If you want to view the TID in the CLI only, run the following command:

```bash
jstack <PID> | grep <TID> -A 30
```

-A 30 indicates that 30 lines are displayed.

----End

### 6.12 HBase Startup Failure and the "FileNotFoundException" Error Is Recorded in RegionServer Logs

**Background**

HBase fails to start and the RegionServer stays in the Concerning state.

**Cause Analysis**

1. Check the RegionServer log (`/var/log/Bigdata/hbase/rs/hbase-omm-XXXout`). It is found that the following information is printed:

```java
org.apache.hadoop.hbase.regionserver.HRegionServer.abort(HRegionServer.java:2123)
java.io.FileNotFoundException: File does not exist: hdfs://hacluster/kylin/kylin_metadata/coprocessor/kylin-coprocessor-1.6.0-SNAPSHOT-0.jar
```

---End

---End
2. Run the hdfs command on the client. It is found that the following file does not exist:

hdfs://hacluster/kylin/kylin_metadata/coprocessor/kylin-coprocessor-1.6.0-SNAPSHOT-0.jar

3. When configuring the coprocessor for HBase, make sure that the path of the corresponding JAR package is correct. Otherwise, HBase cannot be started.

Solution

Use the Apache Kylin engine to interconnect with MRS and make sure that the JAR file of the Kylin engine exists.

6.13 The Number of RegionServers Displayed on the Native Page Is Greater Than the Actual Number After HBase Is Started

Background

After HBase is started, the number of RegionServers displayed on the HMaster native page is greater than the actual number.

The HMaster native page shows that four RegionServers are online, as shown in the following figure.

![RegionServers Displayed on the Native Page](image)

Cause Analysis

As shown in the following figure, the hostname of the node in the third row is controller-192-168-1-3 and that of the fourth row is eth0. The two carry the same information reported by RegionServer. Then, log in to the corresponding
nodes to check the /etc/hosts file. It is found that the same IP address is configured for the two hostnames. For details, see the following figure:

![hosts file content]

### Solution

Log in to the node where RegionServer resides and modify the /etc/hosts file. Make sure that the same IP address can correspond to only one hostname.

### 6.14 The RegionServer Instance Is Abnormal and in the Concerning State.

#### Background

HBase fails to start and the RegionServer stays in the Concerning state.

#### Cause Analysis

Check the running log (/var/log/Bigdata/hbase/rs/hbase-omm-XXX.log) of the abnormal RegionServer instance. It is found that the following information is displayed: **ClockOutOfSyncException..., Reported time is too far out of sync with master.**

```
2017-09-18 11:16:23,636 | FATAL | regionserver21302 | Master rejected startup because clock is out of sync | org.apache.hadoop.hbase.regionserver.HRegionServer.reportForDuty(HRegionServer.java:2059)
org.apache.hadoop.hbase.ClockOutOfSyncException: org.apache.hadoop.hbase.ClockOutOfSyncException: Server nl-bi-fi-datanode-24-65,21302,1505726180086 has been rejected; Reported time is too far out of sync with master. Time difference of 152109ms > max allowed of 30000ms at org.apache.hadoop.hbase.master.ServerManager.checkClockSkew(ServerManager.java:354)
```

...
This log indicates that the time difference between the abnormal RegionServer instance and the HMaster instance is greater than the allowed time difference 30s (specified by the `hbase.regionserver.maxclockskew` parameter and the default value is 30000ms). As a result, the RegionServer instance is abnormal.

**Solution**

Adjust the node time to ensure that the time difference between nodes is less than 30s.

### 6.15 HBase Startup Failure in the Newly Installed Cluster

**Background**

HBase of a newly installed cluster fails to start up. The RegionServer log contains the following error information:

```
2018-02-24 16:53:03,863 | ERROR | regionserver/host3/187.6.71.69:21302 | Master passed us a different hostname to use; was=host3, but now=187-6-71-69 |
org.apache.hadoop.hbase.regionserver.HRegionServer.handleReportForDutyResponse(HRegionServer.java:1386)
```

**Cause Analysis**

In the `/etc/hosts` file, an IP address maps multiple host names.

**Solution**

1. **Step 1** Modify the mapping between the IP address and host names in the `/etc/host` file.
2. **Step 2** Restart HBase.

----End

### 6.16 HBase Startup Failure Due to the Lost ACL Table Directory

**Symptom**

Cluster HBase fails to start.

**Cause Analysis**

1. Check the HMaster log of HBase. The following error information is displayed:
2. The HBase directory in HDFS is checked, which shows that the ACL table directory is lost.

Browse Directory

<table>
<thead>
<tr>
<th>Permission</th>
<th>Owner</th>
<th>Group</th>
<th>Size</th>
<th>Last Modified</th>
<th>Replication</th>
<th>Block Size</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>drwx-------</td>
<td>hbase</td>
<td>supergroup</td>
<td>0 B</td>
<td>Thu Mar 15 21:30:29 2018</td>
<td>0</td>
<td>0 B</td>
<td>meta</td>
</tr>
<tr>
<td>drwx-------</td>
<td>hbase</td>
<td>supergroup</td>
<td>0 B</td>
<td>Thu Mar 15 21:30:36 2018</td>
<td>0</td>
<td>0 B</td>
<td>namespace</td>
</tr>
</tbody>
</table>

Solution

**Step 1** Stop HBase.

**Step 2** Log in to the HBase client as the hbase user and run the following command.

Example:
```
hadoop03:~ # source /opt/client/bigdata_env
hadoop03:~ # kinit hbase
Password for hbase@HADOOP.COM:
hadoop03:~ # hbase zkcli
```

**Step 3** Delete the ACL table information from the ZooKeeper.

Example:
```
[zk: hadoop01:24002,hadoop02:24002,hadoop03:24002(CONNECTED) 0] deleteall /hbase/table/hbase:acl
[zk: hadoop01:24002,hadoop02:24002,hadoop03:24002(CONNECTED) 0] deleteall /hbase/table-lock/hbase:acl
```

**Step 4** Start HBase.

----End

6.17 HBase Fails to Start After the Cluster Is Powered Off and On

**Symptom**

After the ECS in the cluster is shut down and restarted, HBase fails to start.
Cause Analysis

Check the HMaster run logs. A large number of errors are reported, as shown below:

```
2018-03-26 11:10:54,185 | INFO  | hadoopc1h3,21300,1522031630949_splitLogManager__ChoreService_1 | total tasks = 1 unassigned = 0 tasks={/hbase/splitWAL/WALs%2Fhadoopc1h1%2C21302%252C1520214023667.default.1520584926990=last_update = 1522033841041 last_version = 34255 cur_worker_name = hadoopc1h3,21302,1520943011826 status = in_progress incarnation = 3 resubmits = 3 batch = installed = 1 done = 0 error = 0} | org.apache.hadoop.hbase.master.SplitLogManager$TimeoutMonitor.chore (SplitLogManager.java:745)
2018-03-26 11:11:00,185 | INFO  | hadoopc1h3,21300,1522031630949_splitLogManager__ChoreService_1 | total tasks = 1 unassigned = 0 tasks={/hbase/splitWAL/WALs%2Fhadoopc1h1%2C21302%252C1520214023667.default.1520584926990=last_update = 1522033841041 last_version = 34255 cur_worker_name = hadoopc1h3,21302,1520943011826 status = in_progress incarnation = 3 resubmits = 3 batch = installed = 1 done = 0 error = 0} | org.apache.hadoop.hbase.master.SplitLogManager$TimeoutMonitor.chore (SplitLogManager.java:745)
2018-03-26 11:11:06,185 | INFO  | hadoopc1h3,21300,1522031630949_splitLogManager__ChoreService_1 | total tasks = 1 unassigned = 0 tasks={/hbase/splitWAL/WALs%2Fhadoopc1h1%2C21302%252C1520214023667.default.1520584926990=last_update = 1522033841041 last_version = 34255 cur_worker_name = hadoopc1h3,21302,1520943011826 status = in_progress incarnation = 3 resubmits = 3 batch = installed = 1 done = 0 error = 0} | org.apache.hadoop.hbase.master.SplitLogManager$TimeoutMonitor.chore (SplitLogManager.java:745)
2018-03-26 11:11:10,787 | INFO  | RpcServer.reader=9,bindAddress=hadoopc1h3,port=21300 | Kerberos principal name is hbase/hadoop.hadoop.com@HADOOP.COM | org.apache.hadoop.hbase.ipc.RpcServer$Connection.readPreamble(RpcServer.java:1532)
2018-03-26 11:11:18,185 | INFO  | hadoopc1h3,21300,1522031630949_splitLogManager__ChoreService_1 | total tasks = 1 unassigned = 0 tasks={/hbase/splitWAL/WALs%2Fhadoopc1h1%2C21302%252C1520214023667.default.1520584926990=last_update = 1522033841041 last_version = 34255 cur_worker_name = hadoopc1h3,21302,1520943011826 status = in_progress incarnation = 3 resubmits = 3 batch = installed = 1 done = 0 error = 0} | org.apache.hadoop.hbase.master.SplitLogManager$TimeoutMonitor.chore (SplitLogManager.java:745)
```

The WAL splitting of RegionServer fails when the node is powered on and off.

Solution

**Step 1** Stop HBase.

**Step 2** Run the `hdfs fsck` command to check the health status of the `/hbase/WALs` file.

```
hdfs fsck /hbase/WALs
```

If the following command output is displayed, all files are normal. If any file is abnormal, rectify the fault, and then perform the following operations.

```
The filesystem under path `/hbase/WALs` is HEALTHY
```

**Step 3** Back up the `/hbase/WALs` file.

```
hdfs dfs -mv /hbase/WALs /hbase/WALs_old
```
Step 4 Run the following command to create a new `/hbase/WALs` directory.

```
hdfs dfs -mkdir /hbase/WALs
```

Make sure that the permission on the directory is `hbase:hadoop`.

Step 5 Start HBase.

----End

6.18 HBase Is Abnormal

Issue

HBase is abnormal.

Symptom

Error Message "NotServingRegionException" is displayed when data is imported to hbase.

Cause Analysis

When a block size is greater than 2 GB, a read exception occurs during the seek operation of the HDFS. Full GC occurs when data is frequently written to the RegionServer. As a result, the heartbeat between the HMaster and RegionServer is abnormal, and the HMaster marks the RegionServer as dead, the RegionServer is forcibly restarted. After the restart, the servercrash mechanism is triggered to roll back WALs. Currently, the `splitwal` file has reached 2.1 GB and has only one block. As a result, the HDFS seek operation is abnormal and the `splitwal` file fails. RegionServer detects that the WAL needs to be split and triggers the `splitwal` log mechanism for rollback. As a result, the region on the RegionServer node cannot be brought online, and an error message indicating that the region is not online is displayed when a region on the RegionServer node is queried.

Procedure

Step 1 Log in to MRS Manager.

Step 2 Choose Services > HBase. On the Service Status tab page, click HMaster (Active) on the right of HMaster Web UI. The HBase Web UI page is displayed.

Step 3 On the Procedures tab page, view the node where the problem occurs.

Step 4 Log in to the faulty node as user root and run the `hdfs dfs -ls` command to view all block information.

Step 5 Run the `hdfs dfs -mkdir` command to create a directory for storing problem blocks.

Step 6 Run the `hdfs dfs -mv` command to move the faulty block to the new directory.

----End
Summary and Suggestions

The following two points are for reference:

- If data blocks are corrupted, run the `hdfs fsck /tmp -files -blocks -racks` command to check the health information about the data blocks.
- When a region is being split, the `NotServingRegionException` error is thrown if you perform operations on data.
7.1 Content Recorded in Hive Logs

Audit log

An audit log records at what time a user sends a request to HiveServer and MetaStore from which IP address with what statement.

The following HiveServer audit log shows that at 14:51:22 on February 1, 2016, user_chen sent a show tables request to HiveServer from the 192.168.1.18 IP address.

```
2016-02-01 14:51:22,235 | INFO | HiveServer2-HandlerPool-Thread-37015 | UserChen | user_chen | showtables | Time=2016/02/01 14:51:22 | Operation=exec | Statement=show tables | Result=OK | ResultSize=1 |
```

The following MetaStore audit log shows that at 11:31:15 on January 29, 2016 user hive sent a shutdown request to MetaStore from the 192.168.1.18 IP address.

```
2016-01-29 11:31:15,451 | INFO | pool-6-thread-70648 | ugi=hive/hadoop.hadoop.com | shuttingdown the object store... |
```

Generally, the audit log does not play a role in actual error location. However, the audit log must be checked to solve the following problems:

1. There is no response after a client sends a request. The audit log can be used to check whether the task suspends on the client or server. If the audit log has no related information, the task suspends on the client. If the audit log has related information, view the run log to locate where the program suspends.

2. The audit log can be used to check the number of requests in a specified period of time. You can view the number of requests in a specified period in audit logs.

HiveServer Run Log

HiveServer receives requests from a client (SQL statement), compile and execute the statement (submitted to Yarn or local MapReduce), and interact with
MetaStore to obtain metadata information. The HiveServer run log records a complete SQL execution process.

Generally, if SQL statement running fails, check the HiveServer run log first.

**MetaStore Run Log**

Typically, if the HiveServer run log contains MetaException or MetaStore connection failure, check the MetaStore run log.

**GC Log**

Both HiveServer and MetaStore have GC logs. If GC-related problems occur, view the GC logs to quickly locate the cause. For example, if HiveServer or MetaStore frequently restarts, check its GC log.

### 7.2 Causes of Hive Startup Failure

The most common cause of the Hive startup failure is that the MetaStore instance cannot connect to DBService. You can view the detailed error information in the MetaStore logs. The reasons for the failure to connect to DBService are as follows:

**Possible Cause 1**

DBService does not properly initialize the Hive metabase hivemeta.

**Procedure 1**

1. Run the following commands:
   
   ```
   source /opt/Bigdata/MRS_XXX/install/dbservice/.dbservice_profile
gsql -h 192.168.0.44 (Floating IP address of DB) -p 20051 -d hivemeta -U hive -W HiveUser@
   ```

2. If the interaction interface cannot be properly displayed, database initialization fails. If the following error information is displayed, the hivemeta configuration may be lost in the configuration file of the node where DBService is located.

   ```
   org.postgresql.util.PSQLException: FATAL: no pg_hba.conf entry for host "192.168.0.146", database "HIVEMETA"
   ```

3. Edit `/srv/BigData/dbdata_service/data/pg_hba.conf` by adding `host hivemeta hive 0.0.0.0/0 sha256` to the file.

4. Run the `source /opt/Bigdata/MRS_XXX/install/dbservice/.dbservice_profile` command to configure environment variables.

5. Run `gs_ctl -D $GAUSSDATA reload #` to make new configurations take effect.

   ----End

**Possible Cause 2**

The floating IP address of DBService is incorrect. As a result, the IP address of the MetaStore node fails to connect to or build mutual trust with the floating IP address, causing MetaStore startup failure.
Procedure 2

The floating IP address of DBService must be an IP address that is not used in the same network segment and cannot be pinged before configuration. Modify the floating IP address of DBService.

7.3 "Cannot modify xxx at runtime" Is Reported When the set Command Is Executed in a Security Cluster

Symptom

The following error is reported when running the set command:

```
0: jdbc:hive2://192.168.1.18:21066/> set mapred.job.queue.name=QueueA;
Error: Error while processing statement: Cannot modify mapred.job.queue.name at list of params that are allowed to be modified at runtime (state=42000,code=1)
```

Procedure

Solution 1:

Step 1 Log in to MRS Manager and choose **Services > Hive > Service Configuration**. Set **Type** to **All** and choose **HiveServer > Security**.

Step 2 Add the command parameters to be executed to the **hive.security.authorization.sqlstd.confwhitelist.append** configuration item.

Step 3 Click **Save** and restart **HiveServer**. For details, see the following figure:

---End

Solution 2:
Step 1  Log in to MRS Manager and choose Services > Hive > Service Configuration. Set Type to All and choose HiveServer > Security.

Step 2  Locate hive.security.whitelist.switch and select OFF. Click Save and restart HiveServer.

---End

7.4 How to Specify a Queue When Hive Submits a Task

Symptom

How do I specify a queue when Hive submits a task?

Procedure

Step 1  Before executing the statement, set the task queue by using the following parameters. For example, submit the task to QueueA.

    set mapred.job.queue.name=QueueA;
    select count(*) from rc;

Step 2  After task submission, you can find that the task is submitted to QueueA on the Yarn page.

<table>
<thead>
<tr>
<th>User:</th>
<th>admin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name:</td>
<td>select count(*) from rc(Stage-1)</td>
</tr>
<tr>
<td>Application Type:</td>
<td>MAPREDUCE</td>
</tr>
<tr>
<td>Application Tags:</td>
<td></td>
</tr>
<tr>
<td>YarnApplicationState:</td>
<td>FINISHED</td>
</tr>
<tr>
<td>Queue:</td>
<td>QueueA</td>
</tr>
<tr>
<td>FinalStatus Reported by AM:</td>
<td>SUCCEEDED</td>
</tr>
<tr>
<td>Started:</td>
<td>Thu Mar 03 09:01:58 +0800 2016</td>
</tr>
<tr>
<td>Elapsed:</td>
<td>1mins, 0sec</td>
</tr>
<tr>
<td>Tracking URL:</td>
<td>History</td>
</tr>
<tr>
<td>Log Aggregation Status</td>
<td>Status</td>
</tr>
<tr>
<td>Diagnostics:</td>
<td></td>
</tr>
</tbody>
</table>

NOTE

Queue names are case-sensitive. Both queueA and Queuea are invalid.

---End

7.5 How to Set Map and Reduce Memory on the Client

Symptom

How do I set Map and Reduce memory on the client?
Procedure

Before SQL statement execution, run the set command to set parameters of clients related to Map/Reduce.

The following parameters are related to Map and Reduce memory:

- `set mapreduce.map.memory.mb=4096; //Memory required by each Map task`
- `set mapreduce.map.java.opts=-Xmx3276M; //Maximum memory used by the JVM of each Map task`
- `set mapreduce.reduce.memory.mb=4096; //Memory required by each Reduce task`
- `set mapreduce.reduce.java.opts=-Xmx3276M; //Maximum memory used by the JVM of each Reduce task`
- `set mapred.child.java.opts=-Xms1024M -Xmx3584M; //This parameter is a global parameter, which is used to set Map and Reduce in a unified manner.`

**NOTE**

Parameter settings take effect for the current session only.

### 7.6 How to Specify the Output File Compression Format When Importing a Table

**Symptom**

How do I specify an output file compression format when importing a table?

**Procedure**

Hive supports the following compression formats:

- `org.apache.hadoop.io.compress.BZip2Codec`
- `org.apache.hadoop.io.compress.Lz4Codec`
- `org.apache.hadoop.io.compress.DeflateCodec`
- `org.apache.hadoop.io.compress.SnappyCodec`
- `org.apache.hadoop.io.compress.GzipCodec`

- To compress all tables, choose Services > Hive > Service Configuration on MRS Manager and set the following Hive service parameters:
  - Set `hive.exec.compress.output` to `true`.
  - Set `mapreduce.output.fileoutputformat.compress.codec` to `org.apache.hadoop.io.compress.BZip2Codec`.

  **NOTE**

  The following parameters take effect only when `hive.exec.compress.output` is set to `true`.

- If it needs to be set at the session level, configure the parameters as follows before command execution:

  ```
  set hive.exec.compress.output=true;
  set mapreduce.output.fileoutputformat.compress.codec=org.apache.hadoop.io.compress.SnappyCodec;
  ```
7.7 The desc Table Is Too Long to Be Completely Displayed

**Symptom**

How do I make sure that the description is completely displayed when the desc table is too long?

**Procedure**

**Step 1** When starting the beeline of Hive, set `maxWidth` to **20000**, for example:

```
[root@192-168-1-18 logs]# beeline --maxWidth=20000
```

scan complete in 3ms
Connecting to

Beeline version 1.1.0 by Apache Hive

**Step 2** (Optional) Run the `beeline -help` command to view the client display settings. For details, see the following content:

- `-u <database url>` the JDBC URL to connect to
- `-n <username>` the username to connect as
- `-p <password>` the password to connect as
- `-d <driver class>` the driver class to use
- `-i <init file>` script file for initialization
- `-e <query>` script file that should be executed
- `-hivconf property=value` Use value for given property
- `-color=[true/false]` control whether color is used for display
- `-showHeader=[true/false]` show column names in query results
- `-fastConnect=[true/false]` skip building table/column list for tab-completion
- `-autoCommit=[true/false]` enable/disable automatic transaction commit
- `-verbose=[true/false]` show verbose error messages and debug info
- `-showWarnings=[true/false]` display connection warnings
- `-showNestedErrs=[true/false]` display nested errors
- `-numberFormat=[pattern]` format numbers using DecimalFormat pattern
- `-force=[true/false]` continue running script even after errors
- `-maxWidth=MAXWIDTH` the maximum width of the terminal
- `-maxColumnWidth=MAXCOLWIDTH` the maximum width to use when displaying columns
- `-silent=[true/false]` be more silent
- `-autosave=[true/false]` automatically save preferences
- `-delimiterForDSV=DELIMITER` specify the delimiter for delimiter-separated values output format
- `-isolation=LEVEL` set the transaction isolation level
- `-nullemptystring=[true/false]` set to true to get historic behavior of printing null as empty string
- `-socketTimeOut=n` socket connection timeout interval, in second. The default value is 300.

----End
7.8 Data Inserted After the Partition Column Is Added Is Displayed as NULL

**Symptom**

1. Run the following command to create a table:
   ```sql
   create table test_table(
   col1 string,
   col2 string
   )
   PARTITIONED BY(p1 string)
   STORED AS orc tblproperties('orc.compress'='SNAPPY');
   ```

2. Modify the table structure, add partitions, and insert data.
   ```sql
   alter table test_table add partition(p1='a');
   insert into test_table partition(p1='a') select col1,col2 from temp_table;
   ```

3. Modify the table structure, add columns, and insert data.
   ```sql
   alter table test_table add columns(col3 string);
   insert into test_table partition(p1='a') select col1,col2,col3 from temp_table;
   ```

4. Query data in the `test_table` table. In the returned result, the values in the `col3` column are all NULL.
   ```sql
   select * from test_table where p1='a'
   ```

5. Add a table partition and insert data.
   ```sql
   alter table test_table add partition(p1='b');
   insert into test_table partition(p1='b') select col1,col2,col3 from temp_table;
   ```

6. Query data in the `test_table` table. In the returned result, the value of `col3` is not all NULL.
   ```sql
   select * from test_table where p1='b'
   ```

**Cause Analysis**

RESTRICT is the default option for altering a table. In the RESTRICT mode, only the metadata is changed, while the table's partition structure created before the altering operation remains unchanged. However, new partitions created after the altering operation are changed. Therefore, when values of the old partitions are queried, they are all NULL.

**Procedure**

Add the `cascade` keyword when adding columns, for example:
```sql
alter table test_table add columns(col3 string) cascade;
```

7.9 A Newly Created User Has No Query Permissions

**Symptom**

When a user is created, an error message is displayed indicating that the user does not have permissions to query data.

Error: Error while compiling statement: FAILED: HiveAccessControlException Permission denied: Principal [name=hive, type=USER] does not have following privileges for operation QUERY [[SELECT] on Object [type=TABLE_OR_VIEW, name=default.t1]] (state=42000,code=40000)
Cause Analysis

The newly created user does not have the permission to operate the Hive component.

Solution

Step 1  Log in to MRS Manager and choose System > Manage Role > Create Role.

Step 2  Enter a role name.

Step 3  In the Permission area, select Hive. The Hive administrator permission and the read and write permission for Hive tables are displayed.

Step 4  Select Hive Read Write Privileges. All databases in the Hive column are displayed.

Step 5  Select the permissions required by the role and click OK.

Step 6  On MRS Manager, choose System > Manage User.

Step 7  Locate the row that contains the created user and click Modify in the Operation column.

Step 8  Click Select and Join User Group. To use the Hive service, you must add a Hive group.

Step 9  Click Select and Add Role and select the role created in Step 5.

Step 10  Click OK.

----End
7.10 An Error Is Reported When Executing SQL to Submit a Task to a Specified Queue

**Symptom**

The following error is reported when executing SQL to submit a task to Yarn:

```
Failed to submit application_1475400939788_0033 to YARN: org.apache.hadoop.security.AccessControlException: User newtest cannot submit applications to queue root.QueueA
```

**Cause Analysis**

The current login user does not have the permission to submit the YARN queue.

**Solution**

Grant the submission permission of the specified Yarn queue to the user. On MRS Manager, choose **System > Manage User** and bind a role with the queue submission permission to the user.

---

7.11 An Error Is Reported When The `load data inpath` Command Is Executed

**Symptom**

The following errors are reported when the `load data inpath` command is executed:

- **Error 1:**
  HiveAccessControlException Permission denied. Principal [name=user1, type=USER] does not have following privileges on Object [type=DFS_URI, name=hdfs://hacluster/tmp/input/mapdata] for operation LOAD: [OBJECT OWNERSHIP]

- **Error 2:**
  HiveAccessControlException Permission denied. Principal [name=user1, type=USER] does not have following privileges on Object [type=DFS_URI, name=hdfs://hacluster/tmp/input/mapdata] for operation LOAD: [INSERT, DELETE]

- **Error 3:**
  SemanticException [Error 10028]: Line 1:17 Path is not legal "file:///tmp/input/mapdata": Move from: file:/tmp/input/mapdata to: hdfs://hacluster/user/hive/warehouse/tmp1 is not valid. Please check that values for params "default.fs.name" and "hive.metastore.warehouse.dir" do not conflict.

**Cause Analysis**

The current login user does not have the permission to operate the directory or the file directory format is incorrect.

**Solution**

Hive has the following requirements on the `load data inpath` command:
The file owner must be the user who executes the command.
The current user must have read and write permissions for the file.
The current user must have permissions to execute the directory of the file.
The current user must have the write permission on the directory of the table, because the load operation moves the file to the directory.
The file format must be the same as the storage format specified by the table. For example, if stored as rcf is specified during table creation but the file format is TXT, it is unsatisfied.
The file must be stored in HDFS. Files in the local file system cannot be specified using the file:// form.
The file name cannot start with an underscore (_) or period (.). A file whose name starts with an underscore (_) or period (.) will be ignored.

The following shows permissions required when user test_hive loads data.

```
[root@192-168-1-18 duan]# hdfs dfs -ls /tmp/input2
Found 1 items
-rw-r--r--   3 test_hive hive 6 2016-03-21 14:44 /tmp/input2/input.txt
```

7.12 An Error Is Reported When The load data local inpath Command Is Executed

**Symptom**

The following errors are reported when the load data local inpath command is executed:

- **Error 1:**
  HiveAccessControlException Permission denied. Principal [name=user1, type=USER] does not have following privileges on Object [type=LOCAL_URI, name=file:/tmp/input/mapdata] for operation LOAD : [SELECT, INSERT, DELETE]

- **Error 2:**
  HiveAccessControlException Permission denied. Principal [name=user1, type=USER] does not have following privileges on Object [type=LOCAL_URI, name=file:/tmp/input/mapdata] for operation LOAD : [OBJECT OWNERSHIP]

- **Error 3:**
  SemanticException Line 1:23 Invalid path "/tmp/input/mapdata": No files matching path file:/tmp/input/mapdata

**Cause Analysis**

The current user does not have the permission to operate the directory or the directory does not exist on the node where HiveServer is located.

**Solution**

**NOTE**

Generally, you are not advised to use local files to load data to Hive tables. You are advised to store local files in HDFS and then load data from the cluster.

Hive has the following requirements on the load data local inpath command:
- The file must be stored on the HiveServer node, because all commands are sent to the active HiveServer for execution.
- User omm must have the read permission for the file and read and execution permissions for the directory where the file is located, because the HiveServer process is started by user omm in the OS.
- The file owner must be the user who executes the command.
- The current user must have read and write permissions for the file.
- The file format must be the same as the storage format specified by the table. For example, if stored as rcfile is specified during table creation but the file format is TXT, it is unsatisfied.
- The file name cannot start with an underscore (_) or period (.). A file whose name starts with an underscore (_) or period (.) will be ignored.

7.13 An Error Is Reported When The create external table Command Is Executed

Symptom

The following error is reported when the create external table xx(xx int) stored as textfile location '/tmp/aaa/aaa' command is executed.

```
Permission denied. Principal [name=fantasy, type=USER] does not have following privileges on Object [type=DFS_URI, name=/tmp/aaa/aaa] for operation CREATETABLE : [SELECT, INSERT, DELETE, OBJECT OWNERSHIP] (state=42000,code=40000)
```

Cause Analysis

The current login user does not have the read and write permissions for the directory or its parent directory. When an external table is created, whether the current user is checked for its read and write permissions for the specified directory and its subdirectories and subfiles. If the specified directory does not exist, permissions for the parent directory are checked, and so on. If the check results show that the user has no permissions on any directory, insufficient permission instead of no specified directory is reported.

Solution

Check whether the current user has read and write permissions for the /tmp/aaa/aaa path. If the path does not exist, check whether the user has read and write permissions for its parent directory.

7.14 An Error Is Reported When The dfs -put Command Is Executed On The Beeline Client

Symptom

Run following commands:

```
dfs -put /opt/kv1.txt /tmp/kv1.txt
```
The following error is reported:

Permission denied. Principal [name=admin, type=USER] does not have following privileges
onObject[type=COMMAND_PARAMS,name=[-put, /opt/kv1.txt, /tmp/kv1.txt]] for operation DFS : [ADMIN PRIVILEGE] (state=,code=1)

**Cause Analysis**

The current login user does not have the permissions to run the command.

**Solution**

If the current user has the admin role, run the `set role admin` command to switch to the admin role. If you do not have the admin role, choose System > Manage User on MRS Manager and bind the role to the user.

### 7.15 An Error Is Reported Indicating Insufficient Permissions When the set role admin Command Is Executed

**Symptom**

Run following commands:

```
set role admin
```

The following error is reported:

```
0: jdbc:hive2://192.168.42.26:21066/> set role admin:
Error: Error while processing statement: FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. dmp_B doesn't belong to role admin (state=08S01,code=1)
```

**Cause Analysis**

The current user does not have the permissions of the admin role of Hive.

**Solution**

**Step 1** Log in to MRS Manager, choose System > Manage Role > Create Role, and create a role with the Hive Admin Privilege permission.

**Step 2** On MRS Manager, choose System > Manage User.

**Step 3** In the Operation column of the user, click Modify.

**Step 4** Bind a role with the Hive Admin Privilege permission to the user and click OK.

----End
7.16 An Error Is Reported When a UDF Is Created Using Beeline

Symptom

Run following commands:

```sql
create function fn_test3 as 'test.MyUDF' using jar 'hdfs:///tmp/udf2/MyUDF.jar'
```

The following error is reported:

```
Error: Error while compiling statement: FAILED: HiveAccessControlException Permission denied: Principal [name=admin, type=USER] does not have following privileges for operation CREATEFUNCTION [[ADMIN PRIVILEGE] on Object [type=DATABASE, name=default], [ADMIN PRIVILEGE] on Object [type=FUNCTION, name=default.fn_test3]] (state=42000,code=40000)
```

Cause Analysis

To create a permanent function in Hive, role `admin` is required.

Solution

Run the `set role admin` command before running the statement.

7.17 Hive Status Is Bad

Possible Causes

- The DBService service is not available.
- The HDFS service is not available.
- The ZooKeeper service is not available.
- The LDAP/KrbServer service is not available.
- The MetaStore instance is not available.

Solution

- If the DBService is unavailable, see [ALM-27001 DBService Unavailable](#).
- If the HDFS service is unavailable, see [ALM-14000 HDFS Service Unavailable](#).
- If the ZooKeeper service is unavailable, see [ALM-13000 ZooKeeper Service Unavailable](#).
- If the LDAP/KrbServer service is unavailable, see [ALM-25000 LdapServer Service Unavailable/ALM-25500 KrbServer Service Unavailable](#).
- If the MetaStore instance is unavailable, see [ALM-16004 Hive Service Unavailable](#).
7.18 Hive Service Status Is Partially Healthy

Possible Causes

The HBase service is not available.

Solution

If the HBase service is unavailable, see ALM-19000 HBase Service Unavailable

7.19 Difference Between Hive Service Health Status and Hive Instance Health Status

Symptom

What is the difference between Hive service health status and Hive instance health status?

Solution

The Hive service health status is displayed on the Services page and has four values: Good, Bad, Partially Healthy, and Unknown. It depends not only on Hive service availability but also the service status of other related components. Simple SQL is used to check Hive service availability.

Hive instances consist of HiveServer and MetaStore. Their health status is determined by communications between instances and JMX and can be Good (normal communications), Concerning (abnormal communications), or Unknown (no communications).

7.20 Alarms in Hive and Triggering Scenarios

Hive Alarms

<table>
<thead>
<tr>
<th>Alarm ID</th>
<th>Alarm Severity</th>
<th>Auto Clear</th>
<th>Alarm Name</th>
<th>Alarm Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>16000</td>
<td>Minor</td>
<td>TRUE</td>
<td>Percentage of Sessions Connected to the HiveServer to Maximum Number Allowed Exceeds the Threshold</td>
<td>Fault alarm</td>
</tr>
<tr>
<td>Alarm ID</td>
<td>Alarm Severity</td>
<td>Auto Clear</td>
<td>Alarm Name</td>
<td>Alarm Type</td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>16001</td>
<td>Minor</td>
<td>TRUE</td>
<td>Hive Warehouse Space Usage Exceeds the Threshold</td>
<td>Fault alarm</td>
</tr>
<tr>
<td>16002</td>
<td>Minor</td>
<td>TRUE</td>
<td>The Successful Hive SQL Operations Lower than The Threshold</td>
<td>Fault alarm</td>
</tr>
<tr>
<td>16004</td>
<td>Critical</td>
<td>TRUE</td>
<td>Hive Service Unavailable</td>
<td>Fault alarm</td>
</tr>
</tbody>
</table>

Alarm Triggering Scenarios

- **16000**: An alarm is triggered when the ratio of the number of sessions connected to HiveServer to the allowed total number of sessions exceeds the threshold. For example, if the number of connected sessions is 9, the allowed total number of sessions is 12, and the threshold is 70%, an alarm is triggered, because 9/12 > 70%.
- **16001**: An alarm is triggered when the ratio of HDFS capacities used by Hive to total HDFS capacities allocated to Hive exceeds the threshold. For example, if 500 GB is allocated to Hive, Hive uses 400 GB, and the threshold is 75%, an alarm is triggered, because 400/500 > 75%.
- **16002**: An alarm is triggered when SQL execution success rate is lower than the threshold. If two out of four SQL statements are executed successfully and the threshold is 60%, an alarm is triggered, because 2/4 < 60%.
- **16004**: An alarm is triggered when the health status of the Hive service changes to Bad.

**NOTE**

- You can choose **System > Configure Alarm Threshold** on MRS Manager to set the alarm threshold, alarm severity, and time segment for triggering the alarm.
- Metrics related to Hive running can be viewed on the Hive monitoring interface.

### 7.21 "authentication failed" Is Displayed During Shell Client Connection

**Symptom**

In clusters in security mode, the **beeline** command fails to be executed on the Shell client when the HiveServer service is normal, and the system prompts "authentication failed". The following information is displayed.
Cause Analysis

- The client user does not perform security authentication.
- Kerberos authentication expired.

Solution

Step 1 Log in to the node where the Hive client is installed.

Step 2 Run the `source $client_home/bigdata_env` command.

Run the `klist` command to check whether there is a valid ticket in the local end. The following information shows that the ticket became valid at 14:11:42 on December 24, 2016, and expired at 14:11:40 on December 25, 2016. In the period of time, the ticket was available.

```
klist
Ticket cache: FILE:/tmp/krb5cc_0
Default principal: admin@HADOOP.COM
Valid starting     Expires            Service principal
12/24/16 14:11:42  12/25/16 14:11:40  krbtgt/HADOOP.COM@HADOOP.COM
```

Step 3 Run the `kinit username` command for authentication and log in to the client again.

----End

7.22 Client Displays a Message Indicating ZooKeeper Access Failure

Symptom

In clusters in security mode, when the HiveServer service is normal and SQL is executed by using the JDBC interface to connect to HiveServer, "The ZooKeeper client is AuthFailed" is reported.

```
14/05/19 10:52:00 WARN utils.HAClientUtilDummyWatcher: The ZooKeeper client is AuthFailed
14/05/19 10:52:00 INFO utils.HiveHAClientUtil: Exception thrown while reading data from znode.The possible reason may be connectionless. This is recoverable. Retrying..  
14/05/19 10:52:16 WARN utils.HAClientUtilDummyWatcher: The ZooKeeper client is AuthFailed
14/05/19 10:52:32 WARN utils.HAClientUtilDummyWatcher: The ZooKeeper client is AuthFailed
14/05/19 10:52:32 ERROR st.BasicTestCase: Exception: Could not establish connection to active hiveserver
java.sql.SQLException: Could not establish connection to active hiveserver
```

Or an error is reported stating "Unable to read HiveServer2 configs from ZooKeeper":

```
Exception in thread "main" java.sql.SQLException: org.apache.hive.jdbc.ZooKeeperHiveClientException: Unable to read HiveServer2 configs from ZooKeeper
```
Caused by: org.apache.hive.jdbc.ZooKeeperHiveClientException: Unable to read HiveServer2 configs from ZooKeeper
at org.apache.hive.jdbc.ZooKeeperHiveClientHelper.configureConnParams(ZooKeeperHiveClientHelper.java:100)
at org.apache.hive.jdbc.Utils.configureConnParams(Utils.java:429)
at org.apache.hive.jdbc.HiveConnection.<init>(HiveConnection.java:142)

Caused by: org.apache.zookeeper.KeeperException$ConnectionLossException: KeeperErrorCode = ConnectionLoss for /hiveserver2
at org.apache.zookeeper.KeeperException.create(KeeperException.java:99)
at org.apache.zookeeper.ZooKeeper.getChildren(ZooKeeper.java:2374)
at org.apache.curator.framework.imps.GetChildrenBuilderImpl$3.call(GetChildrenBuilderImpl.java:214)
at org.apache.curator.framework.imps.GetChildrenBuilderImpl$3.call(GetChildrenBuilderImpl.java:203)
at org.apache.curator.RetryLoop.callWithRetry(RetryLoop.java:107)
at org.apache.curator.framework.imps.GetChildrenBuilderImpl.pathInForeground(GetChildrenBuilderImpl.java:200)
at org.apache.curator.framework.imps.GetChildrenBuilderImpl.forPath(GetChildrenBuilderImpl.java:38)

Cause Analysis

- When the client connects to HiveServer, the HiveServer address is automatically obtained from ZooKeeper. If ZooKeeper connection authentication is abnormal, the HiveServer address cannot be obtained from ZooKeeper correctly.

- During ZooKeeper connection authentication, `krb5.conf`, `principal`, `keytab`, and related information must be loaded to the client. Authentication failure causes are as follows:
  - The user.keytab path is incorrectly entered.
  - user.principal is incorrectly entered.
  - The cluster has switched the domain name. However, the old principal is used when the client combines the URL.
  - The client cannot pass Kerberos authentication due to firewall settings. Ports 21730 (TCP), 21731 (TCP/UDP), and 21732 (TCP/UDP) need to be opened for Kerberos.

Solution

**Step 1** Ensure that the user can properly access the user.keytab file in related paths on the client node.

**Step 2** Ensure that the user's user.principal corresponds to the specified keytab file.

Run the `klist -kt keytabpath/user.keytab` command to check the file.

**Step 3** If the cluster has switched the domain name, the principal field used in the URL must be the new domain name.

For example, the default value is `hive/hadoop.hadoop.com@HADOOP.COM`. If the cluster has switched the domain name, the field must be changed accordingly. For example, if the domain name is abc.com, enter `hive/hadoop.abc.com@ABC.COM`. 
Step 4  Ensure that authentication is normal and HiveServer can be connected.

Run the following commands on the client:

```bash
source $client_home/bigdata_env
kinit username
```

Run the `beeline` command on the client to ensure normal running.

---End

7.23 "Invalid function" Is Displayed When a UDF Is Used

**Symptom**

When a UDF is created on the Hive client using Spark, "Error 10011" indicating "invalid function" is reported:

```
Error: Error while compiling statement: FAILED: SemanticException [Error 10011]: Line 1:7 Invalid function 'test_udf' (state=42000,code=10011)
```

The preceding problem occurs when multiple HiveServers use a UDF. For example, if metadata is not synchronized in time when the UDF created on HiveServer2 is used on HiveServer1, the preceding error is reported when clients on HiveServer1 are connected.

**Cause Analysis**

Metadata shared by multiple HiveServers or Hive and Spark is not synchronized, causing memory data inconsistency between different HiveServer instances and invalid UDF.

**Solution**

Synchronize new UDF information to HiveServer and reload the function.

7.24 Hive Service Status Is Unknown

**Cause Analysis**

The Hive service stops.

**Solution**

Restart the Hive service.
7.25 Health Status of a HiveServer or MetaStore Instance Is Unknown

**Symptom**

The health status of a HiveServer or MetaStore instance is unknown.

**Cause Analysis**

The HiveServer or MetaStore instance is stopped.

**Solution**

Restart the HiveServer or MetaStore instance.

7.26 Health Status of a HiveServer or MetaStore Instance Is Concerning

**Symptom**

The health status of the HiveServer or MetaStore instance is **Concerning**.

**Cause Analysis**

The HiveServer or MetaStore instance cannot be normally started. For example, when modifying the MetaStore/HiveServer GC parameter, you can view the startup log of the corresponding process, for example, the `hiveserver.out(hadoop-omm-jar-192-168-1-18.out)` file. The following exception occurs:

Error: Could not find or load main class Xmx2048M

The preceding information indicates that Xmx2048M is used as the startup parameter of the Java process instead of the JVM during the startup of the Java virtual machine. As shown in the following information, the hyphen (-) is deleted mistakenly.

```
```

**Solution**

Check the latest changes to detect incorrect settings.

```
```
7.27 If Text Files Are Compressed Using the ARC4 Algorithm, Garbled Characters Are Returned After the select Query

Symptom

If a Hive query result table is compressed and stored using the ARC4 algorithm, garbled characters are returned after the select * query is conducted in the result table.

Cause Analysis

The default Hive compression format is not ARC4 or output compression is disabled.

Solution

Step 1 If garbled characters are returned after the SETECT query, set the following in Beeline:

```sql
set mapreduce.output.fileoutputformat.compress.codec=org.apache.hadoop.io.encryption.arc4.ARC4BlockCodec;
set hive.exec.compress.output=true;
```

Step 2 Import the table to a new table using block decompression.

```sql
insert overwrite table tbl_result select * from tbl_source;
```

Step 3 Perform the query again.

```sql
select * from tbl_result;
```

7.28 Hive Task Running Fails But Task Re-execution Is Successful

Symptom

When Hive task running fails, an error similar to the following is reported on the client:

```
Error:Invalid OperationHandler:OperationHandler
[opType=EXECUTE_STATEMENT,getNodeIdentifier()=XXX](state=,code=0)
```

However, the MapReduce task that is submitted by the task to Yarn is successfully executed.
Cause Analysis

The cluster where the error occurs has two HiveServer instances. The error in the log of one HiveServer instance is the same as the error (Error: Invalid OperationHandler) reported on the client. In the log of the other HiveServer instance, START_UP information similar to the following is printed when the error occurs, which indicates that the process is killed and restarted during that time. Because the HiveServer instance the task process plans to connect to is killed, it connects to the other healthy one, causing the error.

```
2017-02-15 14:40:11,309 | INFO  | main | STARTUP_MSG: ************************************************************
STARTUP_MSG: Starting HiveServer2
STARTUP_MSG:   host = XXX-120-85-154/XXX.120.85.154
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 1.3.0
```

Solution

Submit the task again and ensure that the HiveServer process is not manually restarted during task execution.

7.29 An Error Is Reported When the SELECT Statement Is Executed

Symptom

When the `select count(*) from XXX` statement is executed, the client reports the error "Error:Error while processing statement :FAILED:Execution Error,return code 2 from...".

return code 2 indicates that the task fails because an error is reported during the execution of the MapReduce task.
Cause Analysis

1. Go to the native Yarn page to check the MapReduce task logs. The check result shows that the error occurs due to unidentified compression mode. The file name suffix is .gz but the stack reports .zlib.

2. Therefore, the HDFS file corresponding to the table that is queried may be incorrect. According to the file name printed in the map log, download the file from HDFS to the local end. The file whose name is suffixed with .gz fails to be decompressed by running the `tar` command because its format is incorrect. Run the `file` command to check the file property. The command output shows that the file is compressed from the FAT system instead of UNIX.
Solution

Delete the file with an incorrect format from the HDFS directory or replace it with a correct one.

7.30 The drop partition Operation Fails When A Large Number of Partitions Exist

Symptom

Exceptions occur during the drop partition operation:

```java
MetaStoreClient lost connection. Attempting to reconnect. 
org.apache.hadoop.hive.metastore.RetryingMetaStoreClient.invoke(RetryingMetaStoreClient.java:187) 
org.apache.thrift.transport.TTransportException 
at org.apache.thrift.transport.TIOStreamTransport.read(TIOStreamTransport.java:132) 
at org.apache.thrift.transport.TTransport.readAll(TTransport.java:86) 
at org.apache.thrift.transport.TSaslTransport.readLength(TSaslTransport.java:376) 
at org.apache.thrift.transport.TSaslTransport.readFrame(TSaslTransport.java:453) 
at org.apache.thrift.transport.TSaslTransport.read(TSaslTransport.java:435) 
```

The MetaStore log shows StackOverflow errors.

```
2017-04-22 01:00:58,834 | ERROR | pool-6-thread-208 | java.lang.StackOverflowError 
at org.datanucleus.store.rdbms.sql.SQLText.toSQL(SQLText.java:330) 
at org.datanucleus.store.rdbms.sql.SQLText.toSQL(SQLText.java:339) 
at org.datanucleus.store.rdbms.sql.SQLText.toSQL(SQLText.java:339) 
at org.datanucleus.store.rdbms.sql.SQLText.toSQL(SQLText.java:339) 
at org.datanucleus.store.rdbms.sql.SQLText.toSQL(SQLText.java:339)
```

Cause Analysis

The processing logic of the drop partition operation is to find all the partitions that meet the conditions, combine them, and delete them together. However, because the number of partitions is too large and the data stack for deleting metadata is deep, StackOverflow errors occur.

Solution

Delete partitions in batches.

7.31 Failed to Start a Local Task

Symptom

1. When operations such as join are performed and the data volume is small, the local task execution fails:
   ```
   jdbc:hive2://10.*.*.*:21066/> select a.name ,b.sex from student a join student1 b on (a.name = b.name); 
   ERROR : Execution failed with exit status: 1 
   ERROR : Obtaining error information 
   ERROR : Task failed! 
   Task ID: 
   Stage-4 
   ... 
   Error: Error while processing statement: FAILED: Execution Error, return code 1 from
   ```
2. **The HiveServer log shows that the local task failed to start.**

   Task ID:
   Stage-4
   Logs:
   | org.apache.hadoop.hive.ql.session.SessionState$LogHelper.printError(SessionState.java:1016)
   at org.apache.hive.service.cli.operation.SQLOperation.toSQLException(Operation.java:339)
   at org.apache.hive.service.cli.operation.SQLOperation.runQuery(SQLOperation.java:169)
   at org.apache.hive.service.cli.operation.SQLOperation.access$200(SQLOperation.java:75)
   at org.apache.hive.service.cli.operation.SQLOperation$1$1.run(SQLOperation.java:245)
   at java.security.AccessController.doPrivileged(Native Method)
   at javax.security.auth.Subject.doAs(Subject.java:422)
   at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1710)
   at org.apache.hive.service.cli.operation.SQLOperation$Worker.runWorker(SQLOperation.java:258)
   at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:511)
   at java.util.concurrent.FutureTask.run(FutureTask.java:266)
   at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1142)
   at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:617)
   at java.lang.Thread.run(Thread.java:745)

3. **The hs_err_pid_****.log file in the HiveServer log directory /var/log/Bigdata/hive/hiveserver contains errors about insufficient memory.**

   # There is insufficient memory for the Java Runtime Environment to continue.
   # Native memory allocation (mmap) failed to map 20776943616 bytes for committing reserved memory.

**Cause Analysis**

When Hive executes join, MapJoin is generated when the data volume is small. When MapJoin is executed, a local task is generated. JVM memory launched by the local task inherits the memory of the parent process.

When multiple joins are executed, multiple local tasks are started. If the host is out of memory, the local tasks fail to start.

**Solution**

**Step 1** On MRS Manager, choose **Services > Hive > Service Configuration** and modify the Hive service configuration parameters.
Step 2  Change the value of `hive.auto.convert.join` in Hive to **false**, save the configuration, and restart the service.

The value change has an adverse impact on service performance. Subsequent operations do not affect service performance.

Step 3  Decrease the value of **Xms** in the `HIVE_GC_OPTS` file of Hive. The minimum value is half of the value of **Xmx**. After the modification, save the configuration and restart the service.

---End

7.32 WebHCat Start Failure

Symptom

WebHCat fails to be started after the host name is changed.

The following error is reported in the WebHCat startup log (`/var/log/Bigdata/hive/webhcat/hive.log`) of the corresponding node.

![Error Log]

Cause Analysis

1. The server account of the MRS WebHCat role involves the host name. If you change the host name after the installation, WebHCat fails to start.

2. The one-to-many or many-to-one association between IP addresses and host names is configured in the `/etc/hosts` file. As a result, the IP address and host name cannot be obtained correctly after the `hostname` and `hostname -i` commands are executed.

Solution

Step 1  Change the host name of the modified node to the host name before the cluster is installed.

Step 2  Check whether the `/etc/hosts` of the node where WebHCat is located is correctly configured.

Step 3  Restart WebHCat.

---End
7.33 Sample Code Error for Hive Secondary Development After Domain Switching

Symptom

In the sample code for Hive secondary development, an error "No rules applied to ****" is reported:

```
Exception in thread "main" java.lang.IllegalArgumentException: No rules applied to ****
```

Cause Analysis

1. The sample code for Hive secondary development loads `core-site.xml` file that is loaded through classload by default. Therefore, you need to put the configuration file to the `classpath` directory of the startup program.

2. If the domain name of the cluster is changed, the `core-site.xml` file will change. You need to download the latest `core-site.xml` file and save it to the `classpath` directory where the sample code for Hive secondary development is located.

Solution

**Step 1** Download the latest client of the Hive cluster to obtain the latest `core-site.xml` file.

**Step 2** Save the `core-site.xml` file to the `classpath` directory where the sample code process for Hive secondary development is located.

----End

7.34 MetaStore Exception Occurs When the Number of DBService Connections Exceeds the Upper Limit

Symptom

By default, the maximum number of connections to DBService is 300. If the number of connections to DBService is greater than 300 when the traffic is heavy, an exception occurs in MetaStore and an error "slots are reserved for non-replication superuser connections" is reported.
com.jolbox.bonecp.BoneCPobtainInternalConnection(BoneCP.java:292)
org.postgresql.util.PSQLException: FATAL: remaining connection slots are reserved for non-replication superuser connections
at org.postgresql.core.v3.ConnectionFactoryImpl.readStartupMessages(ConnectionFactoryImpl.java:643)
at org.postgresql.core.v3.ConnectionFactoryImpl.openConnectionImpl(ConnectionFactoryImpl.java:184)
at org.postgresql.core.ConnectionFactory.openConnection(ConnectionFactory.java:64)
at org.postgresql.jdbc3.AbstractJdbc3Connection.<init>(AbstractJdbc3Connection.java:28)
at org.postgresql.jdbc3g.AbstractJdbc3gConnection.<init>(AbstractJdbc3gConnection.java:20)
at org.postgresql.jdbc4.AbstractJdbc4Connection.<init>(AbstractJdbc4Connection.java:30)
at org.postgresql.jdbc4.jdbc4Connection.<init>(jdbc4Connection.java:22)
at org.postgresql.Driver.makeConnection(Driver.java:392)
at org.postgresql.Driver.connect(Driver.java:266)
at java.sql.DriverManager.getConnection(DriverManager.java:664)
at java.sql.DriverManager.getConnection(DriverManager.java:208)
at com.jolbox.bonecp.BoneCPobtainRawInternalConnection(BoneCP.java:361)
at com.jolbox.bonecp.BoneCPobtainInternalConnection(BoneCP.java:269)
at com.jolbox.bonecp.ConnectionHandle.<init>(ConnectionHandle.java:242)
at com.jolbox.bonecp.PoolWatchThread.fillConnections(PoolWatchThread.java:115)
at com.jolbox.bonecp.PoolWatchThread.run(PoolWatchThread.java:82)
at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1142)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:617)
at java.lang.Thread.run(Thread.java:745)

Cause Analysis

Massive service volume from the customers causes the number of connections to DBService to exceed 300, and the maximum connections to DBService needs to be increased.

Solution

Step 1 Log in to MRS Manager.
Step 2 Choose Services > DBService > Service Configuration and set Type to All. Search for the dbservice.database.max.connections configuration item.
Step 3 Change the value of dbservice.database.max.connections to a proper value. The value cannot exceed 1000.
Step 4 Save the change and restart the affected services or instances.
Step 5 If the number of connections still exceeds the maximum value, check the service code and check whether there are connection leaks.

---End

7.35 Error "Failed to execute session hooks: over max connections" Reported by Beeline

Symptom

The default maximum connections to the HiveServer are 200. When the number of connections exceeds 200, Beeline reports the error "Failed to execute session hooks: over max connections."

beeline> [root@172-27-16-38 c70client]# beeline
Connecting to
jdbc:hive2://129.188.82.38:24002,129.188.82.36:24002,129.188.82.35:24002;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2;asl.qop=auth-conf;auth=KERBEROS;principal=hive/
hadoop.hadoop.com@HADOOP.COM

Debug is true storeKey false useTicketCache true doNotPrompt false ticketCache is null isInitiator true KeyTab is null refreshKrb5Config is false principal is null tryFirstPass is false useFirstPass is false storePass is false clearPass is false

Acquire TGT from Cache
Principal is admin@HADOOP.COM
Commit Succeeded

Error: Failed to execute session hooks: over max connections. (state=,code=0)
Beeline version 1.2.1 by Apache Hive

View the HiveServer log (/var/log/Bigdata/hive/hiveserver/hive.log). It is found that the error "over max connections" is reported.

2018-05-03 04:31:56,728 | WARN | HiveServer2-Handler-Pool: Thread-137 | Error opening session: |
org.apache.hive.service.cli.thrift.ThriftCLIService.OpenSession(ThriftCLIService.java:542)
or.org.apache.hive.service.cli.HiveSQLException: Failed to execute session hooks: over max connections.
at org.apache.hive.service.cli.session.SessionManager.openSession(SessionManager.java:322)
at org.apache.hive.service.cli.CLIService.openSessionWithImpersonation(CLIService.java:189)
at org.apache.hive.service.cli.session.SessionController.openSession(SessionController.java:663)
at org.apache.hive.service.cli.thrift.ThriftCLIService.OpenSession(ThriftCLIService.java:527)
at org.apache.hive.service.cli.thrift.TCLIService$Processor$OpenSession.getResult(TCLIService.java:1257)
at org.apache.hive.service.cli.thrift.TCLIService$Processor$OpenSession.getResult(TCLIService.java:1242)
at org.apache.thrift.TBaseProcessor.process(TBaseProcessor.java:39)
at $TUGIAssumingProcessor.process(HadoopThriftAuthBridge.java:710)
at org.apache.hadoop.hive.thrift.TBaseProcessor$WorkerProcess.run(TBaseProcessor.java:386)
at org.apache.hadoop.hive.thrift.TThreadPoolProcessor.runWorker(TThreadPoolProcessor.java:286)
at org.apache.hadoop.hive.thrift.TThreadPoolExecutor$Worker.run(TThreadPoolExecutor.java:1142)
at java.util.concurrent.ThreadPoolExecutor$Worker.runWorker(ThreadPoolExecutor.java:617)
at java.lang.Thread.run(Thread.java:745)
at org.apache.hadoop.hive.thrift.HadoopThriftAuthBridge$Server.$TUGIAssumingProcessor.process(HadoopThriftAuthBridge.java:710)
at org.apache.hadoop.hive.thrift.TBaseProcessor$WorkerProcess.run(TBaseProcessor.java:386)
at org.apache.hadoop.hive.thrift.TThreadPoolProcessor.runWorker(TThreadPoolProcessor.java:286)
at org.apache.hadoop.hive.thrift.TThreadPoolExecutor$Worker.run(TThreadPoolExecutor.java:1142)
at java.util.concurrent.ThreadPoolExecutor$Worker.runWorker(ThreadPoolExecutor.java:617)
at java.lang.Thread.run(Thread.java:745)

Cause Analysis
1. Massive service volume causes the number of connections to one HiveServer node to exceed 200, and the maximum number of connections to HiveServer needs to be increased.

Solution

Step 1 Log in to MRS Manager, choose Services > Hive > Service Configuration, and set Type to All. Search for the hive.server.session.control.maxconnections configuration item.

Step 2 Change the value of hive.server.session.control.maxconnections to a proper value. The value cannot exceed 1000.

Step 3 Save the change and restart the affected services or instances.

----End
7.36 Beeline Reports the "OutOfMemoryError" Error

Symptom

When a large amount of data is queried on the Beeline client, the message "OutOfMemoryError: Java heap space" is displayed. The detailed error information is as follows:

```
org.apache.thrift.TException: Error in calling method FetchResults
  at org.apache.hive.jdbc.HiveConnection$SynchronizedHandler.invoke(HiveConnection.java:1514)
  at com.sun.proxy.$Proxy4.FetchResults(Unknown Source)
  at org.apache.hive.jdbc.HiveQueryResultSet.next(HiveQueryResultSet.java:358)
  at org.apache.hive.beeline.BufferedRows.<init>(BufferedRows.java:42)
  at org.apache.hive.beeline.BeeLine.print(BeeLine.java:1856)
  at org.apache.hive.beeline.Commands.execute(Commands.java:873)
  at org.apache.hive.beeline.BeeLine.execute(BeeLine.java:821)
  at org.apache.hive.beeline.BeeLine.begin(BeeLine.java:778)
  at org.apache.hive.beeline.BeeLine.mainWithInputRedirection(BeeLine.java:486)
  at org.apache.hive.beeline.BeeLine.main(BeeLine.java:469)
Caused by: java.lang.OutOfMemoryError: Java heap space
  at com.sun.crypto.provider.CipherCore.doFinal(CipherCore.java:959)
  at com.sun.crypto.provider.CipherCore.doFinal(CipherCore.java:824)
  at com.sun.crypto.provider.AESCipher.engineDoFinal(AESCipher.java:436)
  at javax.crypto.Cipher.doFinal(Cipher.java:2223)
  at sun.security.krb5.internal.crypto.dk.AesDkCrypto.decryptCTS(AesDkCrypto.java:414)
  at sun.security.krb5.internal.crypto.dk.AesDkCrypto.decryptRaw(AesDkCrypto.java:291)
  at sun.security.jgss.krb5.CipherHelper.aes256Decrypt(CipherHelper.java:1397)
  at sun.security.jgss.krb5.CipherHelper.decryptData(CipherHelper.java:576)
  at sun.security.jgss.krb5.WrapToken_v2.getData(WrapToken_v2.java:130)
  at sun.security.jgss.krb5.WrapToken_v2.getData(WrapToken_v2.java:105)
  at sun.security.jgss.krb5.Krb5Context.unwrap(Krb5Context.java:1058)
  at sun.security.jgss.GSSContextImpl.unwrap(GSSContextImpl.java:403)
  at com.sun.security.sasl.gsskerb.GssKrb5Base.unwrap(GssKrb5Base.java:77)
  at org.apache.thrift.transport.TSaslTransport$SaslParticipant.unwrap(TSaslTransport.java:559)
  at org.apache.thrift.transport.TSaslTransport.readFrame(TSaslTransport.java:462)
  at org.apache.thrift.transport.TSaslTransport.read(TSaslTransport.java:435)
  at org.apache.thrift.transport.TSaslClientTransport.read(TSaslClientTransport.java:37)
  at org.apache.thrift.transport.TTransport.readAll(TTransport.java:86)
  at org.apache.hadoop.hive.thrift.TFilterTransport.readAll(TFilterTransport.java:62)
  at org.apache.thrift.protocol.TBinaryProtocol.readAll(TBinaryProtocol.java:429)
  at org.apache.thrift.TServiceClient.receiveBase(TServiceClient.java:77)
  at org.apache.hive.service.cli.thrift.TCLIService$Client.recv_FetchResults(TCLIService.java:505)
  at org.apache.hive.service.cli.thrift.TCLIService$Client.FetchResults(TCLIService.java:492)
  at sun.reflect.GeneratedMethodAccessor1.invoke(Unknown Source)
  at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
  at java.lang.reflect.Method.invoke(Method.java:498)
  at org.apache.hive.jdbc.HiveConnection$SynchronizedHandler.invoke(HiveConnection.java:1506)
  at com.sun.proxy.$Proxy4.FetchResults(Unknown Source)
  at org.apache.hive.jdbc.HiveQueryResultSet.next(HiveQueryResultSet.java:358)
Error: Error retrieving next row (state=,code=0)
```

Cause Analysis

- The data volume is excessively large.
- Users use the `select * from table_name` statement for query in the whole table. There is a large amount of data in the table.
- The default startup memory of Beeline is 128 MB. The returned result set is too large during query, overloading Beeline.
Solution

Step 1 Before running `select count(*) from table_name;`, check the amount of data to be queried and determine whether to display data of this magnitude in Beeline.

Step 2 If a certain amount of data needs to be displayed, adjust the JVM parameter of the Hive client. Add `export HIVE_OPTS=-Xmx1024M` (change the value based on service requirements) to `component_env` in the `/Hive` directory of the Hive client. Run the `source` command to obtain the `/bigdata_env` directory on the client.

7.37 Task Execution Fails Because the Input File Number Exceeds the Threshold

Symptom

When Hive performs a query operation, the error message "Job Submission failed with exception 'java.lang.RuntimeException(input file number exceeded the limits in the conf;input file num is: 2380435,max heap memory is: 16892035072,the limit conf is: 500000/4)'" is displayed. The value in the error message varies depending on the actual situation. The error details are as follows:

```
ERROR : Job Submission failed with exception 'java.lang.RuntimeException(input file numbers exceeded the limits in the conf;input file num is: 2380435,max heap memory is: 16892035072,the limit conf is: 500000/4)'
```

Error: Error while processing statement: FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec mr.MapRedTask (state=08S01,code=1)
Cause Analysis

Policy for checking the number of input files before MapReduce job submission: specifies the ratio of maximum number of input files in the submitted MapReduce jobs to maximum HiveServer heap memory, for example, 500000/4 (default value), indicating that each 4 GB heap memory allows 500,000 input files. This error occurs when the number of input files exceeds the limit.

Solution

Step 1 Log in to MRS Manager, choose Services > Hive > Service Configuration, and set Type to All. Search for the hive.mapreduce.input.files2memory configuration item.

Step 2 Change the value of hive.mapreduce.input.files2memory based on the actual memory and task situation.

Step 3 Save the change and restart the affected services or instances.

Step 4 If the problem persists, adjust the GC parameter of the Hive server based on service requirements.

-----End

7.38 Task Execution Fails Because Stack Memory Overflow Is Reported During Task Execution

Symptom

When Hive performs a query operation, the error "Error running child: java.lang.StackOverflowError" is reported. The error details are as follows:

at org.apache.hive.com.esotericsoftware.kryo.util.DefaultClassResolver.readName(DefaultClassResolver.java:127)
at org.apache.hive.com.esotericsoftware.kryo.util.DefaultClassResolver.readClass(DefaultClassResolver.java:115)
at org.apache.hive.com.esotericsoftware.kryo.Kryo.readClass(Kryo.java:656)
at org.apache.hive.com.esotericsoftware.kryo.kryo.readClassAndObject(Kryo.java:767)
at org.apache.hive.com.esotericsoftware.kryo.serializers.collectionSerializer.read(CollectionSerializer.java:112)

Cause Analysis

The error "java.lang.StackOverflowError" indicates the thread stack overflow. This problem occurs if there are excessive calling hierarchies (for example, infinite recursive call) or the thread stack is small.
Solution

Adjust the stack memory in the JVM parameters of the Map and Reduce subprocesses in the MapReduce phase. The involved parameters are `mapreduce.map.java.opts` (adjusting the stack memory of Map) and `mapreduce.reduce.java.opts` (adjusting the stack memory of Reduce). The following uses the `mapreduce.map.java.opts` parameter as an example.

- To increase the Map memory temporarily:
  Run the `set mapreduce.map.java.opts=-Xss8G;` command on the Beeline client. (Change the value as required.)

- To permanently increase the Map memory specified by the `mapreduce.map.memory.mb` and `mapreduce.map.java.opts` parameters:
  a. Log in to MRS Manager, choose Services > Hive > Service Configuration, and set Type to All.
  b. Add the custom parameter `mapreduce.map.java.opts` and its value.
  c. Save the change and restart the affected services or instances.

The modifications take effect after the modifications are saved and the service is restarted. During the restart, the Hive service is unavailable. Pay attention to the execution time window.

7.39 Task Failure Due to Concurrently Writing Data to One Table or Partition

Symptom

When Hive executes an INSERT statement, an error is reported indicating that a file or directory already exists or is cleared in HDFS. The error details are as follows:

```
Caused by: java.io.IOException: File xxx already exists in HDFS
```

Cause Analysis

1. Check the start time and end time of the task based on the HiveServer audit logs.
2. Check whether data is inserted into the same table or partition in the time segment.
3. Hive does not support concurrent data insertion for a table or partition. As a result, multiple tasks perform operations on the same temporary data directory, and one task moves the data of another task, causing task failure.

Solution

The service logic is modified so that data is inserted to the same table or partition in single thread mode.
7.40 Hive Task Failure Due to a Lack of HDFS Directory Permission

Symptom

An error message is displayed, indicating that the user does not have the permission to access the HDFS directory.

```
2019-04-09 17:49:19,845 | ERROR | HiveServer2-Background-Pool: Thread-3160445 | Job Submission failed with exception 'org.apache.hadoop.security.AccessControlException(Permission denied: user=hive_quanxian, access=READ_EXECUTE, inode="/user/hive/warehouse/bigdata.db/

gd_ga_wa_swywswjylzhongao@hive:drwx------
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkAccessAcl(FSPermissionChecker.java:426)
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.check(FSPermissionChecker.java:329)
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkSubAccess(FSPermissionChecker.java:300)
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkPermission(FSPermissionChecker.java:241)
at com.huawei.hadoop.adapter.hadoop.HWAccessControlEnforce.checkPermission(HWAccessControlEnforce.java:69)
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkPermission(FSPermissionChecker.java:190)
at org.apache.hadoop.hdfs.server.namenode.FSDirectory.checkPermission(FSDirectory.java:1910)
at org.apache.hadoop.hdfs.server.namenode.FSDirectory.checkPermission(FSDirectory.java:1894)
at org.apache.hadoop.hdfs.server.namenode.FSDirectory.checkPermission(FSDirectory.java:1894)
at org.apache.hadoop.hdfs.server.namenode.FSNamesystem getContentSummary(FSNamesystem.java:3983)
at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer getContentSummary(NameNodeRpcServer.java:1342)
at org.apache.hadoop.ipc.ProtobufRpcEngine$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine.java:616)
at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:973)
at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2260)
at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2256)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:422)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1781)
at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2254)
')
```

Cause Analysis

1. According to the stack information, the permission on the subdirectory fails to be checked.

```
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkSubAccess(FSPermissionChecker.java:300)
```

2. Check the permission of all files and directories in HDFS. The permission of a directory is 700 (only the file owner can access the directory). It is confirmed that an abnormal directory exists.
Solution

1. Check whether the file is imported manually. If not, delete the file or directory.
2. If the file or directory cannot be deleted, change the file or directory permission to 770.

7.41 Failed to Load Data to a Hive Table

Symptom

After creating a table, a user runs the LOAD command to import data to the table. However, the following problem occurs during the import:

```
> LOAD DATA INPATH '/user/tester1/hive-data/data.txt' INTO TABLE employees_info;
Error: Error while compiling statement: FAILED: SemanticException Unable to load data to destination table.
Error: The file that you are trying to load does not match the file format of the destination table.
(state=42000,code=40000)
```

Cause Analysis

1. After analysis, it is found that the storage format is not specified during table creation. Therefore, the default storage format RCFile is used.
2. However, the data to be imported is in TEXTFILE format.

Solution

There are multiple solutions. If the storage format specified by the table is the same as the format of the data to be imported, you can use a proper method based on the site requirements.

- Method 1:
  You can specify the storage format when creating a table. For example,
  ```
  CREATE TABLE IF NOT EXISTS employees_info(name STRING,age INT) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE
  ```
  indicates that the data to be imported is in TEXTFILE format.
- Method 2:
  The storage format of the imported data must be RCFile instead of TEXTFILE.

7.42 Application Developed Based on the Hive JDBC Code Case Fails to Be Executed

Symptom

After a user develops a service application by referring to the `jdbc-examples` sample project of the Hive component, the application fails to be executed. The application reports the following exception:

```
javax.security.sasl.SaslException: GSS initiate failed
at com.sun.security.sasl.gsskerb.GssKrb5Client.evaluateChallenge(Unknown Source)
at org.apache.thrift.transport.TSaslClientTransport.handleSaslStartMessage(TSaslClientTransport.java:94)
at org.apache.thrift.transport.TSaslTransport.open(TSaslTransport.java:271)
at org.apache.thrift.transport.TSaslClientTransport.open(TSaslClientTransport.java:37)
at org.apache.hadoop.hive.thrift.client.TUGIAssumingTransport1.run(TUGIAssumingTransport.java:52)
at org.apache.hadoop.hive.thrift.client.TUGIAssumingTransport$1.run(TUGIAssumingTransport.java:49)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Unknown Source)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1711)
at org.apache.hadoop.hive.thrift.client.TUGIAssumingTransport.open(TUGIAssumingTransport.java:49)
at org.apache.hive.jdbc.HiveConnection.openTransport(HiveConnection.java:260)
at org.apache.hive.jdbc.HiveConnection.createClient(HiveConnection.java:213)
at org.apache.hive.jdbc.HiveConnection.<init>(HiveConnection.java:49)
at org.apache.hive.jdbc.HiveDriver.connect(HiveDriver.java:105)
at java.sql.DriverManager.getConnection(Unknown Source)
at java.sql.DriverManager.getConnection(Unknown Source)
at com.huawei.bigdata.hive.example.JDBCExample.main(JDBCExample.java:107)
Caused by: org.ietf.jgss.GSSException: No valid credentials provided (Mechanism level: Failed to find any Kerberos tgt)
at sun.security.jgss.krb5.Krb5InitCredential.getInstance(Unknown Source)
at sun.security.jgss.krb5.Krb5MechFactory.getCredentialElement(Unknown Source)
at sun.security.jgss.krb5.Krb5MechFactory.getMechanismContext(Unknown Source)
at sun.security.jgss.GSSManagerImpl.getMechanismContext(Unknown Source)
at sun.security.jgss.GSSManagerImpl.getMechanismContext(Unknown Source)
at sun.security.jgss.GSSContextImpl.initSecContext(Unknown Source)
at sun.security.jgss.GSSContextImpl.initSecContext(Unknown Source)
... 17 common frames omitted

Cause Analysis

1. It is suspected that service interaction is performed before Kerberos authentication is complete.

2. Further analyze the logs. It is found that com.huawei.bigdata.security.LoginUtil - Login success!!!!!!!!!!!!!!! is printed in the logs, but org.apache.hadoop.security.UserGroupInformation: Login successful... is not printed.

   Analyze the source code. It is found that:

   `/* */   @InterfaceAudience.Public
   /* */   @InterfaceStability.Evolving
   /* */   public static synchronized void loginUserFromKeytab(String user, String path)
   /* */     throws IOException
   /* */     {
   /* */       if (!isSecurityEnabled()) {
   /* */         return;
   /* */       }
   /* */       /* */
   /* */   }

3. Analyze isSecurityEnabled() and check whether hadoop.security.authentication is set to kerberos in the configuration.

   This Hive service application is not correctly configured. Therefore, the system determines that Kerberos authentication is not required.

   Analyze the jdbc-examples sample project of the Hive component. This problem does not occur in the sample project because the core-site.xml configuration file exists in the classpath directory of the project and hadoop.security.authentication is set to kerberos in the configuration file.

Solution

To solve this problem, you can use either of the following methods:
● Method 1:
Save the core-site.xml configuration file in the classpath directory by referring to the jdbc-examples sample project of the Hive component.

● Method 2:
In the code, explicitly load the configuration file core-site.xml:

```java
conf = new Configuration();
String userdir = System.getProperty("user.dir") + File.separator + "conf" + File.separator;
conf.addResource(new Path(userdir + "core-site.xml"));
```

● Method 3:
In the code, set hadoop.security.authentication to kerberos:

```java
CONF = new Configuration();
CONF.set("hadoop.security.authentication", "kerberos");
```

### 7.43 HiveServer and HiveHCat Process Faults

**Issue**

The HiveServer and WebHCat processes in the customer cluster are faulty.

**Symptom**

The HiveServer and WebHCat processes on the Master2 node in the MRS cluster are faulty. After the restart, the processes are still faulty.

**Cause Analysis**

On MRS Manager, start the faulty HiveServer process. Log in to the background and search for the error information at the corresponding time point in the hiveserver.out log file. The error information is as follows: `error parsing conf mapred-site.xml` and `Premature end of file`. Restart WebHCat. The same error is reported because the mapred-site.xml file fails to be parsed.

**Procedure**

1. Log in to the Master2 node as user root.
2. Run the `find / -name 'mapred-site.xml'` command to obtain the location of the mapred-site.xml file.
   - The path of HiveServer is `/opt/Bigdata/MRS_2.1.0/1_13_HiveServer/etc/mapred-site.xml`.
   - The path of WebHCat is `/opt/Bigdata/MRS_2.1.0/1_13_WebHCat/etc/mapred-site.xml`.
3. Check whether the mapred-site.xml file is normal. In this case, the configuration file is empty. As a result, the parsing fails.
4. Restore the mapred-site.xml file. Run the `scp` command to copy the configuration file in the corresponding directory on the Master1 node to the corresponding directory on the Master2 node to replace the original file.
5. Run the `chown omm:wheel mapred-site.xml` command to change the owner group and user.
6. On MRS Manager, restart the faulty HiveServer and WebHCat processes.

7.44 An Error Is Reported When MRS Hive Uses Code to Connect to ZooKeeper

**Issue**

An error is reported when MRS Hive connects to ZooKeeper.

**Symptom**

Hive 1.2.1 of the MRS 1.8 cluster is successfully connected to the MRS cluster through the JDBC interface of Hive. However, an error is reported when Hive 2.3.2 of the MRS 1.9.0 cluster is connected to the MRS cluster through the JDBC interface of Hive for computing tasks. An error message as shown in the following figure is displayed:

```
Caused by: org.apache.zookeeper.KeeperException$ConnectionLossException: KeeperErrorCode = ConnectionLoss for /hiveserver2
```

**Cause Analysis**

The authentication method of MRS 1.8 is the same as that of the open source version. The Kerberos authentication file is directly transferred to the Hive URL for authentication. However, in MRS 1.9, this method cannot be used. You need to perform Kerberos authentication, obtain some other configuration information, and then combine the URL.

**Procedure**

For details, see the authentication connection of `hive-examples` in the Huawei MRS 1.9 sample project. For details about the sample address, see the Developer Guide.

**Summary and Suggestions**

Customers develop code based on the official MRS 1.9 samples.

7.45 Configuration Data Connection Timed Out

**Issue**

The interface times out during data connection configuration.
Symptom

When the POST /api-gateway/api/v1/datasource/mapping interface is invoked during data connection configuration, the invoking times out. Log in to the API Gateway background and check the Tomcat log. The Tomcat log shows that the interface returns error code 500, but the API Gateway log does not contain any error information.

Figure 7-1 Tomcat log

Figure 7-2 APIGW log

Cause Analysis

The Hive interface driver is incorrectly configured. As a result, the request times out.

Procedure

Step 1 Log in to the active and standby Master nodes of the cluster, switch to user omm, and modify the following configuration files:
The preceding file path uses the MRS 2.1.0 cluster as an example. Change the path based on the cluster version and Hive component version.

**Step 2** Add the following configuration items to the two configuration files:

```
<property type="advanced" scope="all" classification="MetaDB">
  <name>datanucleus.rdbms.datastoreAdapterClassName</name>
  <value>org.datanucleus.store.rdbms.adapter.PostgreSQLAdapter</value>
</property>
```

**Figure 7-3** Modifying the `/opt/Bigdata/MRS_2.1.0/install/FusionInsight-Hive-3.1.0/adapter/conf/controller/configurations.xml` file
Figure 7-4 Modifying the /opt/Bigdata/om-0.0.1/etc/components/MRS_2.1.0/Hive/configurations.xml file

Step 3 Run the following command on the active Master node to restart the Controller service (skip this step on the standby Master node):

```bash
sh /opt/Bigdata/om-0.0.1/sbin/restart-controller.sh
```

Step 4 Log in to MRS Manager, choose Hive > Service Configuration, and set Type to All. Add a parameter to the custom parameter, click Save Configuration, and select Restart the affected services or instances. Update service configurations. (Restarting Hive will affect services. You are advised to perform this operation during off-peak hours.)
Step 5 Log in to the active and standby Master nodes of the cluster and run the following command to check whether the configuration modification takes effect:

```
cd /opt/Bigdata/MRS_2.1.0;grep -A 1 -rni "datanucleus.rdbms" *
```

Step 6 On the Dashboard page of the corresponding cluster on the MRS console, click Manage next to Data Connection to reconfigure a data connection. (The RDS DB instance must be newly created and has no data.)

Data Connection

<table>
<thead>
<tr>
<th>Component Name</th>
<th>Association Type</th>
<th>Data Connection ...</th>
<th>Data Connection ...</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hive</td>
<td>Hive metadata</td>
<td>RDS MySQL data...</td>
<td></td>
<td>Edit, Delete</td>
</tr>
</tbody>
</table>

Step 7 Migrate data to the new RDS database by referring to Migrating PostgreSQL Data Using DRS.

Step 8 Log in to MRS Manager. On the Services page, click Restart in the Operation column of the Hive service to restart Hive and related components. (Restarting components affects services. You are advised to restart the components during off-peak hours.)
### 7.46 An Error Is Reported When the insert into Statement Is Executed on Hive But the Error Information Is Unclear on the Command Interface

#### Issue

An error is reported when a customer uses MRS Hive to execute an SQL statement.

#### Symptom

When a user uses MRS Hive to execute an SQL statement, the following error message is displayed:

![Error reported when MRS Hive is used to execute SQL statements](image)

#### Cause Analysis

1. Check the HiveServer log. The following error information is displayed at the corresponding time point:
2. If no important information is found in the reported error information but the metadata field is found in the stack, the error may be related to MetaStore.

3. Check the MetaStore log. The following error information is displayed:

The preceding error context indicates that an error is reported during SQL statement execution. The error information contains the following information:

Caused by: org.postgresql.util.PSQLException: ERROR: value too long for type character varying(4000)
The SQL statement fails to be executed because the length of all columns exceeds 4000 bytes. The restriction needs to be modified.
Procedure

**Step 1** Log in to any Master node in the cluster as the root user and run the `su - omm` command to switch to the omm user.

**Step 2** Run the following command to log in to GaussDB:

```
gsql -p 20051 -d hivemeta -U username -W password
```

**Step 3** Run the following command to modify the restriction:

```
alter table PARTITION_PARAMS alter column PARAM_VALUE type varchar(6000);
```

----End
8.1 An Error Is Reported When the Split Value Is Changed in the Spark Application

Issue

An error is reported when the split value is changed in the Spark application.

Symptom

The customer needs to modify the maximum split size to make multiple mappers possible for acceleration. However, an error is reported when the `set Parameter` command is executed to modify the Hive value.

Cause Analysis

- When configuring the `hive.security.whitelist.switch` parameter to enable or disable the whitelist in security mode, you must set the parameter that needs to be run in `hive.security.authorization.sqlstd.confwhitelist` file.
The default whitelist does not contain the `mapred.max.split.size` parameter. Therefore, the system displays a message indicating that the maximum split size modification is not allowed.

**Procedure**

**Step 1** Log in to MRS Manager, and choose **Services > Hive > Service Configuration**.

**Step 2** Set **Type** to **All**, search **hive.security.authorization.sqlstd.confwhitelist.append**, and add `mapred.max.split.size` to **hive.security.authorization.sqlstd.confwhitelist.append**. For details, see **Using Hive from Scratch**.

**Step 3** Restart the Hive component after the modification.

**Step 4** Run the `set mapred.max.split.size=1000000` command. If no error is reported, the modification is successful.

---End

### 8.2 An Error Is Reported When Spark Is Used

**Issue**

When Spark is used, the cluster fails to run.

**Symptom**

When Spark is used, the cluster fails to run.
Cause Analysis

- Invalid characters are added during command execution.
- The owner and owner group of the uploaded JAR file is incorrect.

Procedure

Step 1 Run `/bin/spark-submit --class cn.interf.Test --master yarn-client /opt/client/Spark/spark1-1.0-SNAPSHOT.jar` to check whether invalid characters are imported.

Step 2 If they are imported, modify the invalid characters and run the command again.

Step 3 After the command is executed again, other errors occur. Both the owner and the owner group of the JAR file are `root`.

Step 4 Change the owner and the owner group of the JAR file to `omm:wheel`.

----End

8.3 Spark, Hive, and Yarn Unavailable due to Insufficient Disk Capacity

Issue

A critical alarm indicating that Spark, Hive, and Yarn are unavailable is generated due to insufficient disk capacity.

Symptom

The disk capacity of the cluster created by the user is insufficient. As a result, a critical alarm indicating that Spark, Hive, and Yarn are unavailable is generated.

Cause Analysis

Insufficient cluster disk capacity affects HDFS data writing. When the HDFS disk space usage exceeds the threshold, HDFS becomes abnormal. As a result, Spark, Hive, and Yarn are unavailable.

The alarm indicating that Spark, Hive, and Yarn are unavailable is generated due to insufficient disk capacity of the cluster. After the disk capacity is expanded, the alarm is cleared. Therefore, it can be determined that the HDFS function fault is caused by insufficient disk capacity.
Procedure

For details about how to clear the alarm triggered by insufficient disk capacity, see ALM-12017 Insufficient Disk Capacity.

Summary and Suggestions

None

Related Information

For details about how to solve the problem that HDFS disk usage exceeds the threshold, see ALM-14001 HDFS Disk Usage Exceeds the Threshold.

For details about how to add a disk to the MRS cluster, see Adding a New Disk to the MRS Cluster.

8.4 A Spark Job Fails to Run Due to Incorrect JAR File Import

Issue

A Spark job fails to be executed.

Symptom

A Spark job fails to be executed.

Cause Analysis

The imported JAR file is incorrect when the Spark job is executed. As a result, the Spark job fails to be executed.

Procedure

Step 1 Log in to any Master node.

Step 2 Run the `cd /opt/Bigdata/MRS_*install/FusionInsight-Spark-*/spark/examples/jars` command to view the JAR file of the sample program.

NOTE

A JAR file name contains a maximum of 1023 characters and cannot include special characters (;/&>,<$). In addition, it cannot be left blank or full of spaces.

Step 3 Check the executable programs in the OBS bucket. The executable programs can be stored in HDFS or OBS. The paths vary according to file systems.
8.5 A Spark Job Is Pending Due to Insufficient Memory

Issue
Memory is insufficient to submit a Spark job. As a result, the job is in the pending state for a long time or out of memory (OMM) occurs during job running.

Symptom
The job is pending for a long time after being submitted. The following error information is displayed after the job is executed repeatedly:

```
Exception in thread "main" org.apache.spark.SparkException: Job aborted due to stage failure:
Aborting TaskSet 3.0 because task 0 (partition 0) cannot run anywhere due to node and executor blacklist.
Blacklisting behavior can be configured via spark.blacklist.*.
```

Cause Analysis
The memory is insufficient. As a result, the submitted Spark job is in the pending state for a long time.

Procedure

Step 1 Log in to the MRS console, click a cluster name on the Active Clusters page and view the node specifications of the cluster on the Nodes tab page.

Step 2 Add cluster resources owned by the nodemanager process.
1. Log in to MRS Manager and choose Services > Yarn > Service Configuration.
2. Set Type to All, and then search for yarn.nodemanager.resource.memory-mb in the search box to view the value of this parameter. You are advised to set the parameter value to 75% to 90% of the total physical memory of nodes.

Step 3 Modify the Spark service configuration.
1. Log in to MRS Manager and choose Services > Spark > Service Configuration.
2. Set Type to All, and then search for spark.driver.memory and spark.executor.memory in the search box.
Set these parameters to a larger or smaller value based on the complexity and memory requirements of the submitted Spark job. (Generally, the values need to be increased.)
Note

- If a SparkJDBC job is used, search for `SPARK_EXECUTOR_MEMORY` and `SPARK_DRIVER_MEMORY` and modify their values based on the complexity and memory requirements of the submitted Spark job. (Generally, the values need to be increased.)
- If the number of cores needs to be specified, you can search for `spark.driver.cores` and `spark.executor.cores` and modify their values.

**Step 4**

Spark depends on memory for computing. If the preceding modification cannot meet job submission requirements, scale out a cluster by following instructions in [Manually Scaling Out a Cluster](#).

---End

### 8.6 An Error Is Reported During Spark Running

**Issue**

The specified class cannot be found when a Spark job is running.

**Symptom**

The specified class cannot be found when a Spark job is running. The error message is as follows:

```
Exception encountered | org.apache.spark.internal.Logging$class.logError(Logging.scala:91)
```

**Possible Cause**

The default path configured by the user is incorrect.

**Procedure**

**Step 1** Log in to any Master node.

**Step 2** Modify the configuration file in the Spark client directory.

```
Run the /opt/client/Spark/spark/conf/spark-defaults.conf command to open the spark-defaults.conf file and set park.executor.extraClassPath to ${PWD}/*.
```

---End

### 8.7 Driver Displays a Message Indicating That the Executor Memory Exceeds the Threshold

**Symptom**

A Spark task fails to be submitted due to excessive memory usage.
Cause Analysis

The Driver log prints that the applied Executor memory exceeds the cluster limit.

16/02/06 14:11:25 INFO Client: Verifying our application has not requested more than the maximum memory capability of the cluster (6144 MB per container)
16/02/06 14:11:29 ERROR SparkContext: Error initializing SparkContext.
java.lang.IllegalArgumentException: Required executor memory (10240+1024 MB) is above the max threshold (6144 MB) of this cluster!

Spark tasks are submitted to Yarn and the resources used by the Executor to run tasks are managed by Yarn. From the error message, you can see that when a user starts the Executor, 10 GB memory is specified, which exceeds the upper memory limit of each Container set by Yarn. As a result, the task cannot be started.

Solution

Modify the Yarn configuration to increase the restriction on containers. For example, you can adjust parameter `yarn.scheduler.maximum-allocation-mb` to control the resources for starting the Executor. Restart the Yarn service after the modification.

You can modify the configuration as follows:

**Step 1** Verify that a user can log in to MRS Manager.

**Step 2** Choose Services > Yarn > Service Configuration and set Type to All.

**Step 3** In Search, enter `yarn.scheduler.maximum-allocation-mb` to modify the parameter, save the configuration, and then restart the service. See the following figure.

![Figure 8-1 Modifying Yarn service parameters](image)

----End

8.8 In Yarn-cluster mode, Can't get the Kerberos realm is abnormal. (Updated on 20190726)

Symptom

A Spark task fails to be submitted due to authentication exceptions.
Cause Analysis

1. The token used to connect to HDFS cannot be found in the print result of Driver, as shown in the following message:

   16/03/22 20:37:10 WARN Client: Exception encountered while connecting to the server: org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretManager $InvalidToken): token (HDFS_DELEGATION_TOKEN token 192 for admin) can't be found in cache

   16/03/22 20:37:10 WARN Client: Failed to cleanup staging dir .sparkStaging/application_145858192236_0003

   org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretManager $InvalidToken): token (HDFS_DELEGATION_TOKEN token 192 for admin) can't be found in cache

2. On the native Yarn page, ApplicationMaster fails to be started twice and the task exits, as shown in Figure 8-2.

Figure 8-2 ApplicationMaster start failure

3. The following error information is displayed when you check the ApplicationMaster log:

```
Exception in thread "main" java.lang.ExceptionInInitializerError
Caused by: org.apache.spark.SparkException: Unable to load YARN support
Caused by: java.lang.IllegalArgumentException: Can't get Kerberos realm
Caused by: java.lang.reflect.InvocationTargetException
Caused by: KrbException: Cannot locate default realm
Caused by: KrbException: Generic error (description in e-text) (60) - Unable to locate Kerberos realm
org.apache.hadoop.hive.metastore.MetaStoreUtils.newInstance(MetaStoreUtils.java:1410)
... 86 more
Caused by: javax.jdo.JDOFatalInternalException: Unexpected exception caught.
NestedThrowables:java.lang.reflect.InvocationTargetException
... 110 more
```

4. When you execute `./spark-submit --class yourclassname --master yarn-cluster /yourdependencyjars` to submit a task in yarn-cluster mode, the Driver is enabled in the cluster. Because the client's `spark.driver.extraJavaOptions` is loaded, you cannot find the `kdc.conf` file in the corresponding path on the cluster node. As a result, the ApplicationMaster fails to be started.

Solution

When submitting a task on the client, configure the `spark.driver.extraJavaOptions` parameter in the command line. In this way, the `spark.driver.extraJavaOptions` parameter in the `spark-defaults.conf` file under the client path is not automatically loaded during task running. When starting a Spark task, use `--conf` to specify the driver configuration as follows (the quotation marks after `spark.driver.extraJavaOptions` cannot be omitted):
8.9 Starting spark-sql and spark-shell Fails Because the JDK Version Does Not Match

Symptom

The JDK version does not match. As a result, the client fails to start spark-sql and spark-shell.

Cause Analysis

1. The following error information is displayed on the Driver:
   
   Exception Occurs: BadPadding 16/02/22 14:25:38 ERROR Schema: Failed initialising database. Unable to open a test connection to the given database. JDBC url = jdbc:postgresql://ip:port/sparkhivemeta, username = spark. Terminating connection pool (set lazyInit to true if you expect to start your database after your app).

2. When a SparkSQL task is used, DBService needs to be accessed to obtain metadata information. On the client, the ciphertext needs to be decrypted for access. During the use, the user does not follow the process or configure environment variables, and the default JDK version exists in the environment variables of the client. As a result, the decryption program invoked during decryption is abnormal, and the user is locked.

Solution

Step 1  Run the `which java` command to check whether the default Java command is the Java command of the client.

Step 2  If it is not, go to the next step.

   `source $client_path/bigdata_env`

   Run the `kinit username` command and enter the password corresponding to the username to start the task.

--- End
8.10 In Yarn-client Mode, ApplicationMaster Fails to Be Started Twice

Symptom

In Yarn-client mode, ApplicationMaster fails to be started twice.

Cause Analysis

1. Driver exception:

   16/05/11 18:10:56 INFO Client:
   client token: N/A
diagnostics: Application application_1462441251516_0024 failed 2 times due to AM Container for
appattemp_1462441251516_0024_000002 exited with exitCode: 10
For more detailed output, check the application tracking page:https://hdnode5:26001/cluster/app/
application_1462441251516_0024 Then click on links to logs of each attempt.
Diagnostics: Exception from container-launch.
   Container id: container_1462441251516_0024_02_000001

2. The ApplicationMaster log file contains the following error information:

   retrying ... | org.apache.spark.Logging$logError(Logging.scala:75)
   2016-05-12 10:21:24,817 | ERROR | [main] | Failed to connect to driver at 192.168.30.57:23867,
   retrying ... | org.apache.spark.Logging$logError(Logging.scala:75)
   $class.logError(Logging.scala:96)
   org.apache.spark.SparkException: Failed to connect to driver!
at org.apache.spark.deploy.yarn.ApplicationMaster.waitForSparkDriver(ApplicationMaster.scala:426)
   ...
   2016-05-12 10:21:24,925 | INFO | [Thread-1] | Unregistering ApplicationMaster with FAILED (diag
message: Uncaught exception: org.apache.spark.SparkException: Failed to connect to driver!)
   | org.apache.spark.Logging$logInfo(Logging.scala:59)

In Spark-client mode, the task Driver runs on a client node (usually a node outside the cluster). During the startup, the ApplicationMaster process is
started in the cluster. After the process is started, information needs to be
registered with the Driver process. The task can be continued only after the
registration is successful. According to the ApplicationMaster log, the
connection to the Driver fails, which causes the task failure.

Solution

Step 1 Check whether the IP address of the Driver process can be pinged.

Step 2 Start a SparkPI task. Information similar to the following is displayed on the
console:

   16/05/11 18:07:20 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://
   sparkDriver@192.168.1.100:23662]
   16/05/11 18:07:20 INFO Utils: Successfully started service 'sparkDriver' on port 23662.

Step 3 Run the `netstat - anp | grep 23662` command on the node (192.168.1.100 in Step
2) to check whether the port is enabled. The following information indicates that
the port is enabled.

   tcp  0    0  ip:port    :::* LISTEN 107274/java
   tcp  0    0  ip:port    ip:port ESTABLISHED 107274/java

Step 4 Run the `telnet 192.168.1.100 23662` command on the node where
ApplicationMaster is started to check whether the port can be connected. Perform
this operation as both the root and omm users. If information similar to Escape character is ^] is displayed, the connection is normal. If connection refused is displayed, the connection fails and the related port cannot be connected.

If the port is enabled but cannot be connected from other nodes, check the network configuration.

**NOTE**

The port (port 23662 in this example) is randomly selected each time. Therefore, you need to test the port enabled by the task.

----End

### 8.11 When a Spark Task Is Submitted, the Connection to ResourceManager Is Abnormal

**Symptom**

The connection to ResourceManager is abnormal. As a result, Spark tasks fail to be submitted.

**Cause Analysis**

1. The following error information is displayed on the Driver, indicating that port 26004 connecting to the active and standby ResourceManager nodes is rejected:

   15/08/19 18:36:16 INFO RetryInvocationHandler: Exception while invoking getClusterMetrics of class ApplicationClientProtocolPBCClientImpl over 33 after 1 fail over attempts. Trying to fail over after sleeping for 17448ms.
   
   java.net.ConnectException: Call From ip0 to ip1:26004 failed on connection exception: java.net.ConnectException: Connection refused.
   
   INFO RetryInvocationHandler: Exception while invoking getClusterMetrics of class ApplicationClientProtocolPBCClientImpl over 32 after 2 fail over attempts. Trying to fail over after sleeping for 16233ms.
   
   java.net.ConnectException: Call From ip0 to ip2:26004 failed on connection exception: java.net.ConnectException: Connection refused;

2. On MRS Manager, check whether ResourceManager is running properly, as shown in **Figure 8-3**. If Yarn is faulty or an unknown exception occurs on a Yarn service instance, ResourceManager of the cluster may be abnormal.
3. Check whether the client is the latest one in the cluster.
   Check whether the ResourceManager instance has been migrated in the cluster. (Uninstall a ResourceManager instance and add it back to other nodes.)
4. On MRS Manager, click Audit to view audit logs and check whether related operations are recorded.
   Run the ping command to check whether the IP address can be pinged.

**Solution**

- If ResourceManager is abnormal, see the Yarn-related sections to rectify the fault.
- If the client is not the latest, download the client again.
- If the IP address cannot be pinged, contact network management personnel to check the network.

### 8.12 DAYU Fails to Schedule Spark Jobs

**Issue**

DAYU fails to schedule jobs, and a message is displayed indicating that data in the `/thriftserver/active_thriftserver` directory fails to be read.

**Symptom**

DAYU fails to schedule jobs, and a message is displayed indicating that data in the `/thriftserver/active_thriftserver` directory fails to be read.

The error information is as follows: Can not get JDBC Connection, due to KeeperErrorCode = NoNode for /thriftserver/active_thriftserver

**Cause Analysis**

When DAYU submits a Spark job, the JDBC mode of Spark is invoked. Spark starts a ThriftServer process for the client to provide JDBC connections. During the
startup, JDBCServer creates the `active_thriftserver` subdirectory in the `/thriftserver` directory of ZooKeeper, and registers related connection information. If the connection information cannot be read, the JDBC connection is abnormal.

**Procedure**

Check whether the ZooKeeper directory contains the target directory and registration information.

**Step 1** Log in to any Master node as the **root** user and initialize environment variables.

```
source /opt/client/bigdata_env
```

**Step 2** Run the `zkCli.sh -server 'ZookeeperIp:2181'` command to log in to ZooKeeper.

**Step 3** Run the `ls /thriftserver` command to check whether the `active_thriftserver` directory exists.

- If the `active_thriftserver` directory exists, run the `get /thriftserver/active_thriftserver` command to check whether the registered configuration information exists in the directory.
  - If the registered configuration information exists, contact HUAWEI CLOUD technical support.
  - If no registered configuration information exists, go to **Step 4**.
- If the `active_thriftserver` directory does not exist, go to **Step 4**.

**Step 4** Log in to **MRS Manager** and check whether the active/standby status of the Spark JDBCServer instance is unknown.

- If yes, go to **Step 5**.
- If no, contact HUAWEI CLOUD technical support.

**Step 5** Restart the two JDBCServer instances. Check whether the status of the active and standby instances is normal and whether the target directory and data exist in ZooKeeper. If yes, the job is restored. If the instance status is not restored, contact HUAWEI CLOUD technical support.

----End

**Summary and Suggestions**

None

**Related Information**

None
9 Using Kafka

9.1 An Error Is Reported When Kafka Is Run to Obtain a Topic

Issue

An Error is reported when Kafka is run to obtain a topic.

Symptom

An error is reported when the Kafka is run to obtain topics. The error information is as follows:

ERROR org.apache.kafka.common.errors.InvalidReplicationFactorException: Replication factor: 2 larger than available brokers: 0.

Possible Cause

The variable for obtaining the ZooKeeper address is incorrect due to special characters.

Procedure

Step 1  Log in to any Master node.

Step 2  Run the `cat /opt/client/Kafka/kafka/config/server.properties | grep '^zookeeper.connect ='` command to check the variable of the Zookeeper address.

Step 3  Run Kafka again to obtain the topic. Do not add any character to the variables obtained in Step 2.

----End
9.2 How Do I Use Python3.x to Connect to Kafka in a Security Cluster?

**Issue**

How do I connect to a Kafka cluster with Kerberos authentication enabled in the Python3.x environment?

**Symptom**

No operation guide is provided for the user to connect to the Kafka cluster with Kerberos authentication enabled in the Python3.x environment.

**Cause Analysis**

The MRS Developer Guide does not contain instructions in Kafka application development guide in the Python environment.

**Procedure**

**Step 1** Log in to the Master node and run the following command to configure the HUAWEI CLOUD EulerOS image source:

```
wget http://mirrors.myhuaweicloud.com/repo/mirrors_source.sh && sh mirrors_source.sh
```

**Step 2** Run the following commands to compile Python3.x:

```
yum groupinstall "Development tools" -y
yum -y install zlib zlib-devel
yum -y install bzip2 bzip2-devel
yum -y install ncurses ncurses-devel
yum -y install readline readline-devel
yum -y install openssl openssl-devel
yum -y install openssl-static
yum -y install sqlite sqlite-devel
yum -y install gdbm gdbm-devel
yum -y install tk tk-devel
yum -y install libffi libffi-devel
```

**Step 3** After the compilation is successful, run the following command to download and decompress the tgz package of Python3.x:

```
wget https://www.python.org/ftp/python/3.6.7/Python-3.6.7.tgz
```
```bash
tar -zxvf Python-3.6.7.tgz

cd Python-3.6.7

NOTE

You can also download the tgz package of Python3.x from the Python official website. Python-3.6.X is recommended. In version 3.7, the take function of RDD cannot be used.

Step 4 Run the following commands to configure Python3.x, and compile and install it in the /opt/Bigdata/python3 directory:

```
./configure --prefix=/opt/Bigdata/python3 --enable-shared CFLAGS=-fPIC
make & & make install
```

NOTE

The installation directory can be customized.

Step 5 Run the following commands to configure Python3.x variables:

```
echo "/opt/Bigdata/python3/lib" >> /etc/ld.so.conf
ldconfig
ln -s /opt/Bigdata/python3/bin/python3 /usr/bin/python3
ln -s /opt/Bigdata/python3/bin/pip3 /usr/bin/pip3
```

NOTE

The variable directory must be the same as the installation directory specified in Step 4.

Step 6 After the configuration is successful, run the following command to install Kafka in the Python3.x environment:

```
cp /usr/include/gssapi/* /home/omm/kerberos/include/gssapi/
pip3 install kafka-python
pip3 install gssapi
```

Step 7 After the installation is successful, run the source /opt/client/bigdata_env command to log in to the MRS client.

```
NOTE

The username and password for logging in to kinit kafka are those for logging in to MRS Manager.
```

Step 8 Run the Python3.x script.

Sample script:
```
producer:
from kafka import KafkaProducer
producer = KafkaProducer(bootstrap_servers=["broker_ip:21007"],
security_protocol="SASL_PLAINTEXT",
sasl_mechanism="GSSAPI",
sasl_kerberos_service_name="kafka",
sasl_kerberos_domain_name="hadoop.hadoop.com")
for _ in range(100):
    response = producer.send("test-topic", b"testmessage")
```
9.3 Flume Normally Connects to Kafka But Fails to Send Messages

**Symptom**

An MRS cluster is installed, and ZooKeeper, Flume, and Kafka are installed in the cluster.

Flume fails to send data to Kafka.

**Possible Causes**

1. The Kafka service is abnormal.
2. The IP address for Flume to connect to Kafka is incorrect.
3. The size of the message sent from Flume to Kafka exceeds the upper limit.

**Cause Analysis**

The possible reasons why Flume fails to send data to Kafka may be related to Flume or Kafka.

1. On the MRS Manager page, choose **Services > Kafka**. The status of Kafka is **Good** and the monitoring indicators show no errors.
2. Check the Flume log. The log contains `MessageSizeTooLargeException` information, as shown in the following:

```
2016-02-26 14:55:19,126 | WARN | [SinkRunner-PollingRunner-DefaultSinkProcessor] | Produce request with correlation id 349829 failed due to [LOG,7]:
kafka.common.MessageSizeTooLargeException | kafka.utils.Logging$class.warn(Logging.scala:83)
```

The exception shows that the size of data written to Kafka by Flume exceeds the maximum message size specified by Kafka.

3. Check the maximum message size specified by Kafka on MRS Manager. Choose **Services > Kafka > Service Configuration**. On the page that is displayed, set **Type** to **All**. All Kafka configurations are displayed. Enter **message.max.bytes** in the **Search** text box to search.

As shown in Figure 9-2, the maximum size of messages that can be received by the Kafka server of MRS is 1000012 bytes = 977 KB by default.
**Solution**

After confirmation with the customer, data sent by Flume contains messages over 1 MB. Adjust parameters on Kafka to enable the messages to be written to Kafka.

**Step 1** Set `message.max.bytes` to a value that is larger than the current maximum size of the message to be written so that Kafka can receive all messages.

**Figure 9-3** Modifying `message.max.bytes`

![Figure 9-3 Modifying message.max.bytes](image)

*Figure 9-3* is only an example. You can adjust it based on the site requirements.

**Step 2** Set `replica.fetch.max.bytes` to a value that is equal to or larger than the value of `message.max.bytes` so that replicas of partitions on different Brokers can be synchronized to all messages.

Choose Services > Kafka > Service Configuration. On the page that is displayed, set Type to All. All Kafka configurations are displayed. Enter `replica.fetch.max.bytes` in the Search text box to search.

Change the parameter values based on **Figure 9-4**.

**Figure 9-4** Modifying `replica.fetch.max.bytes`

![Figure 9-4 Modifying replica.fetch.max.bytes](image)

**Step 3** Click Save Configuration and restart the Kafka service to make Kafka configurations take effect.
Figure 9-5 Saving configurations

**Step 4** Set `fetch.message.max.bytes` to a value that is equal to or larger than the value of `message.max.bytes` for Consumer service applications to ensure that Consumers can consume all messages.

----End

### 9.4 Producer Fails to Send Data and "NullPointerException" Is Thrown

**Symptom**

An MRS cluster is installed, and ZooKeeper and Kafka are installed in the cluster. When Producer sends data to Kafka, the client throws "NullPointerException."

**Possible Causes**

1. The Kafka service is abnormal.
2. The Jass and Keytab files configured on the Producer client are incorrect.

**Cause Analysis**

The possible reasons why Producer fails to send data to Kafka may be related to Producer or Kafka.

1. On the MRS Manager page, choose **Services > Kafka**. The status of Kafka is **Good** and the monitoring indicators show no errors.
2. Check the Producer client logs. The "NullPointerException" error information is displayed, as shown in Figure 9-7.

Alternatively, the log contains only "NullPointerException" information and no stack information. The problem is caused by JDK self-protection. If much information is printed for the same stack, the JDK self-protection is triggered and stack information is no longer printed. See Figure 9-8.

3. Check the Producer client logs. The error information "Failed to configure SaslClientAuthenticator" is displayed, as shown in Figure 9-9.
4. KafkaChannel fails to be created due to authentication failure. The KafkaChannel obtained through the channel (key) is empty and "NullPointerException" is excessively printed. The preceding log shows that authentication failure is due to an empty password. The reason why the password is empty may be that the password does not match the username.

5. Check the Jaas and Keytab files. The principal is set to `stream` in the Jaas file.

Figure 9-10 Checking the Jaas file

```java
KafkaClient {
    com.sun.security.auth.module.Krb5LoginModule required
debug=false
keyTab="/opt/client/user.keytab"
useTicketCache=false
storeKey=true
principal="stream@HADOOP.COM"
useKeyTab=true;
};
```

The principal parameter is set to `zmk_kafka` in the `user.keytab` file.

Figure 9-11 Viewing the user.keytab file

```
[root@Pri-5-148-6 client]# klist -kt user.keytab
Keytab name: FILE:user.keytab
KRBVNO  Timestamp         Principal
------------------------------------------
  1 12/19/16 16:28:17  zmk_kafka@HADOOP.COM
  1 12/19/16 16:28:17  zmk_kafka@HADOOP.COM
```

The principal in the Jaas file does not correspond to that in the `user.keytab` file.

Solution

**Step 1** After confirmation with the customer, the customer's application automatically and scheduled updates the Jaas file. However, when two different processes update the Jaas file, one process writes a correct principal while the other process writes an incorrect one. As a result, the application is abnormal sometimes.

**Step 2** Modify the Jaas file to ensure that the used principal exists in the Keytab file.
9.5 Producer Fails to Send Data and Throws
"TOPIC_AUTHORIZATION_FAILED"

Symptom
An MRS cluster is installed, and ZooKeeper and Kafka are installed in the cluster. When Producer sends data to Kafka, the client throws "TOPIC_AUTHORIZATION_FAILED."

Possible Causes
1. The Kafka service is abnormal.
2. The Producer client adopts non-security access and access is disabled on the server.
3. The Producer client adopts non-security access and ACL is set for Kafka topics.

Cause Analysis
The possible reasons why Producer fails to send data to Kafka may be related to Producer or Kafka.
1. On the MRS Manager page, choose Services > Kafka. The status of Kafka is Good and the monitoring indicators show no errors.
2. Check the Producer client logs. The logs contain the error information "TOPIC_AUTHORIZATION_FAILED."

```
[root@10-10-144-2 client]# kafka-console-producer.sh --broker-list 10.5.144.2:9092 --topic test
1
[2017-01-24 16:58:36,671] WARN Error while fetching metadata with correlation id 0 : (org.apache.kafka.clients.NetworkClient)
```

Producer accesses Kafka using port 9092, which is a non-security port.

3. On MRS Manager, choose **Services > Kafka > Service Configuration** to view the current Kafka cluster configuration. It is found that the customized configuration `allow.everyone.if.no.acl.found=false` is not configured.

4. If ACL is set to **false**, port 9092 cannot be used for access.

5. Check the Producer client logs. The logs contain the error information "TOPIC_AUTHORIZATION_FAILED."

```
[root@10-10-144-2 client]# kafka-console-producer.sh --broker-list 10.5.144.2:21005 --topic test_acl
1
```

Producer accesses Kafka using port 9092, which is a non-security port.

6. Run the client command to check the ACL permission of the topic.

```
[root@10-10-144-2 client]# kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:24002/kafka --list --topic topic_acl
Current ACLs for resource `Topic:topic_acl`:
User:test_user has Allow permission for operations: Describe from hosts: *
User:test_user has Allow permission for operations: Write from hosts: *
```

If ACL is set for the topic, port 9092 cannot be used for access.

7. Check the Producer client logs. The logs contain the error information "TOPIC_AUTHORIZATION_FAILED."

```
[root@10-10-144-2 client]# kafka-console-producer.sh --broker-list 10.5.144.2:21007 --topic acl
1
```

Producer uses port 21007 to access Kafka.

8. Run the client command `klist` to query the current authenticated user.

```
[root@10-10-144-2 client]# klist
Ticket cache: FILE:/tmp/krb5cc_0
Default principal: test@HADOOP.COM
```

Valid starting  Expires Service principal
01/25/17 11:06:48 01/26/17 11:06:45 krbtgt/HADOOP.COM@HADOOP.COM

The current authentication user is **test**.

9. Run the client command to check the ACL permission of the topic.

```
[root@10-10-144-2 client]# kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:24002/kafka --list --topic acl
Current ACLs for resource `Topic:topic_acl`:
User:test_user has Allow permission for operations: Describe from hosts: *
User:test_user has Allow permission for operations: Write from hosts: *
```

The current authentication user is **test**.
After ACL is set for the topic, user **test_user** has Producer permission. User **test** has no permission to perform Producer operations.

For details about the solution, see **Step 2**.

10. Log in to Kafka Broker using SSH.

Run the `cd /var/log/Bigdata/kafka/broker` command to go to the log directory.

Check the **kafka-authorizer.log** log. It shows that the user does not belong to the **kafka** or **kafkaadmin** group.

```plaintext
2017-01-25 13:26:33,648 | WARN  | [kafka-request-handler-0] | The user is not belong to kafka or kafkaadmin group, authorize failed! | kafka.authorizer.logger (SimpleAclAuthorizer.scala:170)
```

For details about the solution, see **Step 3**.

### Solution

**Step 1** Set `allow.everyone.if.no.acl.found` to `true` and restart the Kafka service.

**Step 2** Use the account with permission for login.

The following provides an example:

**kinit test_user**

Alternatively, grant the user with related permission.

---

**NOTICE**

This operation must be performed by the Kafka administrator (belonging to the **kafkaadmin** group).

The following provides an example:

```bash
kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:24002/kafka --topic topic_acl --producer --add --allow-principal User:test
```

**Step 3** Add the user to the **kafka** or **kafkaadmin** group.

----End

### 9.6 Producer Occasionally Fails to Send Data and the Log Displays "Too many open files in system"

**Symptom**

When Producer sends data to Kafka, it is found that the client fails to send data.
Figure 9-13 Producer fails to send data.

Possible Causes
1. The Kafka service is abnormal.
2. The network is abnormal.
3. The Kafka topic is abnormal.

Cause Analysis
1. On the MRS Manager page, choose Services > Kafka. The status of Kafka is Good and the monitoring indicators show no errors.

Figure 9-14 Kafka service status

2. View the error topic information in the SparkStreaming log.
   Run the Kafka commands to obtain the topic assignment information and copy synchronization information, and check the return result.

   *kafka-topics.sh --describe --zookeeper <zk_host:port/chroot>*

   As shown in Figure 9-15, the topic status is normal. All partitions have normal leader information.
3. Run the `telnet` command to check whether the Kafka can be connected.

```
telnet Kafka service IP address Kafka service port
```

If telnet fails, check the network security group and ACL.

4. Log in to Kafka Broker using SSH.

Run the `cd /var/log/Bigdata/kafka/broker` command to go to the log directory.

Check on `server.log` indicates that the error message is displayed in the log shown in the following figure.

```
Figure 9-16 Log exception
```

5. Output of the `lsdf` command used to check the handle usage of the Kafka process on the current node shows that the number of handles used by the Kafka process reaches 470,000.

```
Figure 9-17 Handles
```
6. Check on service codes indicates that the Producer object is frequently created and therefore is not closed normally.

Solution

**Step 1** Stop the current application to ensure that the number of handles on the server does not increase sharply, which affects the normal running of services.

**Step 2** Optimize the application code to resolve the handle leakage problem.

Suggestion: Use one Producer object globally. After the use is complete, call the Close interface to close the handle.

----End

9.7 Consumer Is Initialized Successfully, But the Specified Topic Message Cannot Be Obtained from Kafka

**Symptom**

An MRS cluster is installed, and ZooKeeper, Flume, Kafka, Storm, and Spark are installed in the cluster.

The customer cannot consume any data using Storm, Spark, Flume or self-programmed Consumer code to consume messages of the specified Kafka topic.

**Possible Causes**

1. The Kafka service is abnormal.
2. The IP address for ZooKeeper connection is incorrectly set.
3. "ConsumerRebalanceFailedException" is thrown.
4. "ClosedChannelException" caused by network problems is thrown.

**Cause Analysis**

Storm, Spark, Flume or user-defined Consumer code can be called Consumer.

1. On the MRS Manager page, choose Services > Kafka. The status of Kafka is Good and the monitoring indicators show no errors.
2. Check whether data can be normally consumed through the Kafka client.

Suppose the client has been installed in the `/opt/client` directory, **test** is the topic name to be consumed, and the IP address of ZooKeeper is 192.168.234.231.

```
    cd /opt/client
    source bigdata_env
    kinit admin
    kafka-topics.sh --zookeeper 192.168.234.231:2181/kafka --describe --topic test
    kafka-console-consumer.sh --topic test --zookeeper 192.168.234.231:2181/kafka --from-beginning
```

If data can be consumed, the cluster service is running properly.

3. Check Consumer configurations. The IP address for connecting to ZooKeeper is incorrect.

- **Flume**
  
  ```
  server.sources.Source02.type=org.apache.flume.source.kafka.KafkaSource
  server.sources.Source02.zookeeperConnect=192.168.234.231:2181
  server.sources.Source02.topic = test
  server.sources.Source02.groupId = test_01
  ```

- **Spark**
  
  ```
  val zkQuorum = "192.168.234.231:2181"
  ```

- **Storm**
  
  ```
  BrokerHosts brokerHosts = new ZKHosts("192.168.234.231:2181");
  ```

- **Consumer API**
  
  ```
  zookeeper.connect="192.168.234.231:2181"
  ```

On MRS Manager, the root path of ZNode where Kafka is stored on ZooKeeper is `/kafka`, which is differentiated from the open source. The address for Kafka to connect to ZooKeeper is **192.168.234.231:kafka**. However, the address for Consumer to connect to ZooKeeper is **192.168.234.231:2181**. Therefore, topic information about Kafka cannot be correctly obtained.

For details about the solution, see **Step 1**.
   $ZKSessionExpireListener@34b41dfe] | org.I0Itec.zkclient.ZkEventThread.run(ZkEventThread.java:77)
   kafka.common.ConsumerRebalanceFailedException: pc-zjqbetl86-1454482884879-2ec95ed3 can't rebalance after 4 retries
   at kafka.consumer.ZookeeperConsumerConnector
   $ZKRebalancerListener.syncedRebalance(ZookeeperConsumerConnector.scala:633)
   at kafka.consumer.ZookeeperConsumerConnector
   $ZKSessionExpireListener.handleNewSession(ZookeeperConsumerConnector.scala:487)
   at org.I0Itec.zkclient.ZkClient$4.run(ZkClient.java:472)
   at org.I0Itec.zkclient.ZkEventThread.run(ZkEventThread.java:71)

The exception shows that the current Consumer does not complete rebalance within the specified retry times. As a result, Kafka Topic-Partition is not allocated to Consumer and Consumer cannot consume messages.

For details about the solution, see Step 3.

5. Check Consumer logs. The error message "Fetching topic metadata with correlation id 0 for topics [Set(test)] from broker [id:26,host: 192-168-234-231,port:9092] failed" and "ClosedChannelException" are printed.

   [2016-03-04 03:33:53,047] INFO Fetching metadata from broker id:26,host: 192-168-234-231,port:9092 with correlation id 0 for 1 topic(s) Set(test) (kafka.client.ClientUtils$)
   java.nio.channels.ClosedChannelException
   at kafka.producer.SyncProducer.liftedTree1$1(SyncProducer.scala:73)
   at kafka.producer.SyncProducer.kafka$producer$SyncProducer$$doSend(SyncProducer.scala:72)
   at kafka.producer.SyncProducer.send(SyncProducer.scala:113)
   at kafka.client.ClientUtils$.fetchTopicMetadata(ClientUtils.scala:58)
   at kafka.client.ClientUtils$.fetchTopicMetadata(ClientUtils.scala:93)
   at kafka.consumer.ConsumerFetcherManager$LeaderFinderThread.doWork(ConsumerFetcherManager.scala:66)
   at kafka.util.ShutdownableThread.run(ShutdownableThread.scala:60)

The exception shows that the current Consumer cannot obtain metadata from the Kafka Broker 192-168-234-231 node and cannot connect to the correct Broker for obtaining messages.

6. Check the network conditions. If the network is normal, check whether mapping between the host and the IP address is configured.
   - Linux
     Run the `cat /etc/hosts` command.
Solution

**Step 1** The IP address for connecting to ZooKeeper is incorrectly configured.

**Step 2** Change the IP address for connecting to ZooKeeper in the Consumer configuration and make it consistent with MRS configuration.

- Flume
  ```java
  server.sources.Source02.type=org.apache.flume.source.kafka.KafkaSource
  server.sources.Source02.zookeeperConnect=192.168.234.231:2181/kafka
  server.sources.Source02.topic = test
  server.sources.Source02.groupId = test_01
  ```
- Spark
  ```scala
  val zkQuorum = "192.168.234.231:2181/kafka"
  ```
- Storm
  ```java
  BrokerHosts brokerHosts = new ZKHosts("192.168.234.231:2181/kafka");
  ```
- Consumer API
  ```java
  zookeeper.connect="192.168.234.231:2181/kafka"
  ```

**Step 3** Rebalance is abnormal.

Multiple Consumers in the same consumer group are successively started and consume data of multiple partitions at the same time, load balancing is performed for Consumers when consumers are fewer than partitions.

The temporary node where the Consumer is stored on ZooKeeper determines read/write permission of which partition of which topic the Consumer has. The path is `/consumers/consumer-group-xxx/owners/topic-xxx/x.`
After the load balancing is triggered, the original Consumer will be recalculated and release occupied partitions, which takes a while. Therefore, new Consumers may fail to preempt the partitions.

Table 9-1 Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Function</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rebalance.max.retries</td>
<td>Maximum number of rebalance retries</td>
<td>4</td>
</tr>
<tr>
<td>rebalance.backoff.ms</td>
<td>Interval for each rebalance retry</td>
<td>2000</td>
</tr>
<tr>
<td>zookeeper.session.timeout.ms</td>
<td>Maximum time allowed to create a session with ZooKeeper</td>
<td>15000</td>
</tr>
</tbody>
</table>

Set the preceding parameters to higher values. The following is for your reference:

- `zookeeper.session.timeout.ms = 45000`
- `rebalance.max.retries = 10`
- `rebalance.backoff.ms = 5000`

Parameter setting must comply with the following rule:

\[
\text{rebalance.max.retries} \times \text{rebalance.backoff.ms} > \text{zookeeper.session.timeout.ms}
\]

**Step 4** The network is abnormal.

In the **hosts** file, mapping between the host name and IP address is not configured. As a result, information cannot be obtained when using the host name for access.

**Step 5** Add the host name to the **hosts** file and make it correspond to the IP address.

- **Linux**

**Figure 9-21** Example 3

- **Windows**
9.8 Consumer Fails to Consume Data and Remains in the Waiting State

Symptom

An MRS cluster is installed, and ZooKeeper and Kafka are installed in the cluster. When the Consumer consumes data from Kafka, the client stays in the Waiting state.

Possible Causes

1. The Kafka service is abnormal.
2. The Consumer client adopts non-security access and access is disabled on the server.
3. The Consumer client adopts non-security access and ACL is set for Kafka topics.

Cause Analysis

The possible reasons why the Consumer fails to consume data from Kafka may be related to the Consumer or Kafka.

1. On the MRS Manager page, choose Services > Kafka. The status of Kafka is Good and the monitoring indicators show no errors.
2. Check the Consumer client log. It is found that the information about the frequent connections and disconnections to the Broker node is printed, as shown in the following output.

```
[root@10-10-144-2 client]# kafka-console-consumer.sh --topic test --zookeeper 10.5.144.2:2181/kafka --from-beginning
```

```
[2017-03-07 09:22:00,658] INFO Fetching metadata from broker BrokerEndPoint(1,10.5.144.2,9092) with correlation id 26 for 1 topic(s) Set(test) (kafka.client.ClientUtils$)
[2017-03-07 09:22:00,659] INFO Connected to 10.5.144.2:9092 for producing (kafka.producer.SyncProducer)
[2017-03-07 09:22:00,659] INFO Disconnecting from 10.5.144.2:9092 (kafka.producer.SyncProducer)
```

The Consumer accesses Kafka using port 21005, which is a non-security port.

3. On MRS Manager, choose Services > Kafka > Service Configuration to view the current Kafka cluster configuration. The customized parameter `allow.everyone.if.no.acl.found=false` is not configured.

4. If ACL is set to false, port 9092 cannot be used for access.

5. Check the Consumer client log. It is found that the information about the frequent connections and disconnections to the Broker node is printed, as shown in the following output.

```
[root@10-10-144-2 client]# kafka-console-consumer.sh --topic acl --zookeeper 10.5.144.2:2181/kafka --from-beginning
```

```
[2017-03-07 09:49:16,992] INFO Fetching metadata from broker BrokerEndPoint(2,10.5.144.3,9092) with correlation id 16 for 1 topic(s) Set(topic_acl) (kafka.client.ClientUtils$)
[2017-03-07 09:49:16,993] INFO Connected to 10.5.144.3:9092 for producing (kafka.producer.SyncProducer)
```

The Consumer accesses Kafka using port 21005, which is a non-security port.

6. Run the client command to check the ACL permission of the topic.

```
[root@10-10-144-2 client]# kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:2181/kafka --list --topic acl
```

Current ACLs for resource 'Topic:topic_acl':

---

**Figure 9-23** The Kafka service status is normal.
User: test_user has Allow permission for operations: Describe from hosts: *
User: test_user has Allow permission for operations: Write from hosts: *

If ACL is set for the topic, port 9092 cannot be used for access.

7. The following information is printed in the Consumer client log:

```
[root@10-10-144-2 client]# kafka-console-consumer.sh --topic topic_acl --bootstrap-server 10.5.144.2:21007 --consumer.config /opt/client/Kafka/kafka/config/consumer.properties --from-beginning --new-consumer
```

[2017-03-07 10:19:18,478] INFO Kafka version : 0.9.0.0 (org.apache.kafka.common.utils.AppInfoParser)

The Consumer uses port 21007 to access Kafka.

8. Run the client command `klist` to query the current authenticated user.

```
[root@10-10-144-2 client]# klist
Ticket cache: FILE:/tmp/krb5cc_0
Default principal: test@HADOOP.COM
Valid starting     Expires            Service principal
01/25/17 11:06:48  01/26/17 11:06:45  krbtgt/HADOOP.COM@HADOOP.COM
```

The current authentication user is **test**.

9. Run the client command to check the ACL permission of the topic.

```
[root@10-10-144-2 client]# kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:24002/kafka --list --topic topic_acl
Current ACLs for resource `Topic:topic_acl`:
User: test_user has Allow permission for operations: Describe from hosts: *
User: test_user has Allow permission for operations: Write from hosts: *
User: test_user has Allow permission for operations: Read from hosts: *
```

If ACL is set for the topic, user **test** does not have the permission to perform the Consumer operation.

For details about the solution, see **Step 2**.

10. Log in to Kafka Broker using SSH.

    Run the `cd /var/log/Bigdata/kafka/broker` command to go to the log directory.

    Check the `kafka-authorizer.log` log. It shows that the user does not belong to the **kafka** or **kafkaadmin** group.

    2017-01-25 13:26:33,648 | WARN  | [kafka-request-handler-0] | The user is not belongs to kafka or kafkaadmin group, authorize failed! | kafka.authorizer.logger (SimpleAclAuthorizer.scala:170)

For details about the solution, see **Step 3**.

### Solution

**Step 1** Set `allow.everyone.if.no.acl.found` to **true** and restart the Kafka service.

**Step 2** Use the account with permission for login.

The following provides an example:

```
kinit test_user
```

Alternatively, grant the user with related permission.
This operation must be performed by the Kafka administrator (belonging to the `kafkaadmin` group).

The following provides an example:

```
kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:24002/kafka --topic topic_acl --consumer --add --allow-principal User:test --group test
```

```
[root@10-10-144-2 client]# kafka-acls.sh --authorizer-properties zookeeper.connect=8.5.144.2:24002/kafka --list --topic topic_acl
Current ACLs for resource `Topic:topic_acl`:
  User:test_user has Allow permission for operations: Describe from hosts: *
  User:test_user has Allow permission for operations: Write from hosts: *
  User:test has Allow permission for operations: Describe from hosts: *
  User:test has Allow permission for operations: Write from hosts: *
  User:test has Allow permission for operations: Read from hosts: *
```

**Step 3** Add the user to the `kafka` or `kafkaadmin` group.

--- End

### 9.9 SparkStreaming Fails to Consume Kafka Messages, and the Message "Error getting partition metadata" Is Displayed

**Symptom**

When SparkStreaming is used to consume messages of a specified topic in Kafka, data cannot be obtained from Kafka. The message "Error getting partition metadata" is displayed.

```
Exception in thread "main" org.apache.spark.SparkException: Error getting partition metadata for 'testtopic'. Does the topic exist?
org.apache.spark.streaming.kafka.KafkaCluster$$anonfun$checkErrors$1.apply(KafkaCluster.scala:366)
org.apache.spark.streaming.kafka.KafkaCluster$$anonfun$checkErrors$1.apply(KafkaCluster.scala:366)
scala.util.Either.fold(Either.scala:97)
org.apache.spark.streaming.kafka.KafkaUtils$.createDirectStream(KafkaUtils.scala:422)
com.huawei.bigdata.spark.examples.FemaleInfoCollectionPrint$.main(FemaleInfoCollectionPrint.scala:45)
com.huawei.bigdata.spark.examples.FemaleInfoCollectionPrint.main(FemaleInfoCollectionPrint.scala)
sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
java.lang.reflect.Method.invoke(Method.java:498)
org.apache.spark.deploy.SparkSubmit$.org$apache$spark$deploy$SparkSubmit$$runMain$762$
org.apache.spark.deploy.SparkSubmit$.doRunMain$1(SparkSubmit.scala:183)
org.apache.spark.deploy.SparkSubmit$.submit(SparkSubmit.scala:208)
org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:123)
org.apache.spark.deploy.SparkSubmit.main(SparkSubmit.scala)
```

**Possible Causes**

1. The Kafka service is abnormal.
2. The Consumer client adopts non-security access and access is disabled on the server.
3. The Consumer client adopts non-security access and ACL is set for Kafka topics.

Cause Analysis

1. On the MRS Manager page, choose **Services > Kafka**. The status of Kafka is **Good** and the monitoring indicators show no errors.

![Kafka service status](image)

**Figure 9-24 Kafka service status**

2. On MRS Manager, choose **Services > Kafka > Service Configuration** to view the current Kafka cluster configuration. It is found that `allow.everyone.if.no.acl.found` is not configured or is set to `false`.
3. If it is set to `false`, the Kafka non-secure port 21005 cannot be used for access.
4. Run the client command to check the ACL permission of the topic.
   ```bash
   [root@10-10-144-2 client]# kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:2181/ kafka --list --topic topic_acl
   Current ACLs for resource 'Topic:topic_acl':
   User:test_user has Allow permission for operations: Describe from hosts: *
   User:test_user has Allow permission for operations: Write from hosts: *
   ```
   If an ACL is configured for a topic, the Kafka non-secure port 21005 cannot be used to access the topic.

Solution

**Step 1** Add the customized configuration `allow.everyone.if.no.acl.found` or change its value to `true` and restart the Kafka service.

**Step 2** Delete the ACL configured for the topic.

The following provides an example:
kinit test_user

NOTICE

This operation must be performed by the Kafka administrator (belonging to the kafkaadmin group).

The following provides an example:

```bash
kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:2181/kafka --remove --allow-principal User:test_user --producer --topic topic_acl
```

```bash
kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:2181/kafka --remove --allow-principal User:test_user --consumer --topic topic_acl --group test
```

----End

9.10 Consumer Fails to Consume Data in a Newly Created Cluster, and the Message "GROUPCOORDINATOR_NOTAVAILABLE" Is Displayed

**Symptom**

A Kafka cluster is created, and two Broker nodes are deployed. The Kafka client can be used for production but cannot be used for consumption. The Consumer fails to consume data, and the message "GROUPCOORDINATOR_NOTAVAILABLE" is displayed. The key log is as follows:

```
2018-05-12 10:58:42,561 | INFO | [kafka-request-handler-3] | [GroupCoordinator 2]: Preparing to restabilize group DemoConsumer with old generation 118 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
2018-05-12 10:59:13,562 | INFO | [executor-Heartbeat] | [GroupCoordinator 2]: Preparing to restabilize group DemoConsumer with old generation 119 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
```

**Possible Causes**

The __consumer_offsets__ cannot be created.

**Cause Analysis**

1. The log indicates a large number of __consumer_offsets__ creation failures.
2. Check on the cluster indicates that the Broker number is 2.
3. The __consumer_offset__ topic file requires that the number of copies is 3. Therefore, the creation fails.
Solution

Expand the cluster to at least three streaming core nodes or perform the following steps to modify service configuration parameters:

**Step 1** Log in to MRS Manager and choose Services > Kafka > Service Configuration.

**Step 2** Set Type to All, search for and change the values of offsets.topic.replication.factor and transaction.state.log.replication.factor to 2.

**Step 3** Click Save Configuration and select Restart the affected services or instances. Click OK to restart the services.

----End

9.11 SparkStreaming Fails to Consume Kafka Messages, and the Message "Couldn't find leader offsets" Is Displayed

**Symptom**

When SparkStreaming is used to consume messages of a specified topic in Kafka, data cannot be obtained from Kafka. The following error message is displayed: Couldn't find leader offsets.

**Possible Causes**

- The Kafka service is abnormal.
- The network is abnormal.
- The Kafka topic is abnormal.

**Cause Analysis**

**Step 1** On the MRS Manager page, choose Services > Kafka. The status of Kafka is Good and the monitoring indicators show no errors.
Step 2  View the error topic information in the SparkStreaming log.

Run the Kafka commands to obtain the topic assignment information and copy synchronization information, and check the return result.

```
kafka-topics.sh --describe --zookeeper <zk_host:port/chroot> --topic <topic name>
```

If information in the following figure is displayed, the topic is normal. All partitions have normal leader information.

**Figure 9-25** Topic distribution information and copy synchronization information

Step 3  Check whether the network connection between the client and Kafka cluster is normal. If no, contact the network team to rectify the fault.

Step 4  Log in to Kafka Broker using SSH.
Run the `cd /var/log/Bigdata/kafka/broker` command to go to the log directory.

Check on `server.log` indicates that the error message is displayed in the log shown in the following figure.

```
java.lang.OutOfMemoryError: Direct buffer memory
at java.nio.Bits.reserveMemory(Bits.java:694)
at java.nio.DirectByteBuffer.<init>(DirectByteBuffer.java:123)
at java.nio.ByteBuffer.allocateDirect(ByteBuffer.java:311)
at sun.nio.ch.Util.getTemporaryDirectBuffer(Util.java:241)
at sun.nio.ch.IOUtil.read(IOUtil.java:195)
at sun.nio.ch.SocketChannelImpl.read(SocketChannelImpl.java:380)
at org.apache.kafka.common.network.PlaintextTransportLayer.read(PlaintextTransportLayer.java:110)
```

Step 5 On MRS Manager, choose Services > Kafka > Service Configuration to view the current Kafka cluster configuration. The value of `-XX:MaxDirectMemorySize` in KAFKA_JVM_PERFORMANCE_OPTS is 1G.

Step 6 If the direct memory is too small, an error is reported. Once the direct memory overflows, the node cannot process new requests. As a result, other nodes or clients fail to access the node due to timeout.

-----End

Solution

Step 1 Log in to MRS Manager and choose Services > Kafka > Service Configuration.

Step 2 Set Type to All, and search for and change the value of KAFKA_JVM_PERFORMANCE_OPTS.

Step 3 Click Save Configuration and select Restart the affected services or instances. Click OK to restart the services.

-----End

9.12 Consumer Fails to Consume Data and the Message " SchemaException: Error reading field 'brokers'" Is Displayed

Symptom

When a Consumer consumes messages of a specified topic in Kafka, the Consumer cannot obtain data from Kafka. The following error message is displayed:

```
org.apache.kafka.common.protocol.types.SchemaException: Error reading field 'brokers': Error reading field 'host': Error reading string of length 28271, only 593 bytes available.
```

Exception in thread "Thread-0" org.apache.kafka.common.protocol.types.SchemaException: Error reading field 'brokers': Error reading field 'host': Error reading string of length 28271, only 593 bytes available
at org.apache.kafka.common.protocol.types.Schema.read(Schema.java:73)
Possible Causes

The JAR versions of the client and server are inconsistent.

Solution

Modify the Kafka JAR package in the Consumer application to ensure that it is the same as that on the server.

9.13 Checking Whether Data Consumed by a Customer Is Lost

Symptom

A Customer saves the consumed data to the database and finds that the data is inconsistent with the production data. Therefore, it is suspected that some of Kafka’s consumed data is lost.

Possible Causes

- The customer code is incorrect.
- An exception occurs when Kafka production data is written.
- The Kafka consumption data is abnormal.

Solution

Check Kafka.

Step 1 Observe the changes of the written and consumed offset through consumer-groups.sh. (Produce a certain number of messages, and consume these messages on the client to observe the changes of the offset.)
Step 2 Create a consumption group, use the client to consume messages, and view the consumed messages.

```
new-consumer:

kafka-console-consumer.sh --topic <topic name> --bootstrap-server <IP1:PORT, IP2:PORT,...> --new-consumer --consumer.config <config file>
```

----End

Check the customer code.

Step 1 Check whether an error is reported when the offset is submitted on the client.

Step 2 If no error is reported, add a printing message to the API that is consumed, and print only the key to view the lost data.

----End

9.14 Failed to Start a Component Due to Account Lock

Symptom

In a new cluster, Kafka fails to be started. Authentication failure causes startup failure.

```
/home/omm/kerberos/bin/kinit -k -t /opt/huawei/Bigdata/etc/2_15_Broker/kafka.keytab kafka/hadoop.hadoop.com -c /opt/huawei/Bigdata/etc/2_15_Broker/11846 failed.
export key tab file for kafka/hadoop.hadoop.com failed.export and check keytab failed.errMsg=}} for Broker #192.168.1.92@192-168-1-92.
[2015-07-11 02:34:33] RoleInstance started failure for ROLE[name: Broker].
[2015-07-11 02:34:34] Failed to complete the instances start operation. Current operation entities: [Broker #192.168.1.92@192-168-1-92], Failure entites: [Broker #192.168.1.92@192-168-1-92].Operation Failed.Failed to complete the instances start operation. Current operation entities: [Broker#192.168.1.92@192-168-1-92], Failure entites: [Broker #192.168.1.92@192-168-1-92].
```

Cause Analysis

Check the Kerberos log `/var/log/Bigdata/kerberos/krb5kdc.log`. It is found that IP addresses outside of the cluster uses the `kafka` account for connections, causing multiple authentication failures. As a result, the `kafka` account is locked.

```
Jul 11 02:49:16 192-168-1-91 krb5kd[1863](info): AS_REQ (2 etypes {18 17}) 192.168.1.93: NEEDED_PREAUTH: kafka/hadoop.hadoop.com@HADOOP.COM for krbtgt/HADOOP.COM@HADOOP.COM, Additional pre-authentication required
Jul 11 02:49:16 192-168-1-91 krb5kd[1863](info): preauth (encrypted_timestamp) verify failure: Decrypt integrity check failed
```

Solution

Log in to a node outside the cluster (for example, 192.168.1.93 in the cause analysis example) and disable Kafka authentication. Wait 5 minutes for the account to be unlocked.
9.15 Kafka Broker Reports Abnormal Processes and the Log Shows "IllegalArgumentException"

Symptom

The Process Fault alarm is reported on MRS Manager. Check whether the faulty process is Kafka Broker.

Possible Causes

Broker configuration is abnormal.

Cause Analysis

1. On MRS Manager, click Alarms to obtain the host information.
2. Log in to Kafka Broker using SSH. Run the `cd /var/log/Bigdata/kafka/broker` command to go to the log directory.

   Check the server.log file. It is found that the "IllegalArgumentException" exception is thrown in the following log stating

   "java.lang.IllegalArgumentException: requirement failed: replica.fetch.max.bytes should be equal or greater than message.max.bytes."

   java.lang(IllegalArgumentException: requirement failed: replica.fetch.max.bytes should be equal or greater than message.max.bytes
   at scala.Predef$.require(Predef.scala:233)
   at kafka.server.KafkaConfig.validateValues(KafkaConfig.scala:959)
   at kafka.server.KafkaConfig.<init>(KafkaConfig.scala:944)
   at kafka.server.KafkaConfig$.fromProps(KafkaConfig.scala:701)
   at kafka.server.KafkaConfig$.fromProps(KafkaConfig.scala:698)
   at kafka.server.KafkaServerStartable$.fromProps(KafkaServerStartable.scala:28)
   at kafka.Kafka$.main(Kafka.scala:60)
   at kafka.Kafka.main(Kafka.scala)

Kafka requires that `replica.fetch.max.bytes` be equal to or greater than `message.max.bytes`.

3. Choose Services > Kafka > Service Configuration. On the Kafka configuration page that is displayed, set Type to All to display all Kafka configurations. Search for `message.max.bytes` and `message.max.bytes`. Search for `replica.fetch.max.bytes`. It is found that its value is less than that of `message.max.bytes`.

Solution

**Step 1** Log in to MRS Manager and choose Services > Kafka > Service Configuration. The Kafka configuration page is displayed.

**Step 2** Set Type to All, search for and modify the `replica.fetch.max.bytes` parameter to ensure that its value is greater than or equal to that of `message.max.bytes`. In this way, replicas of partitions on different Brokers can be synchronized to all messages.

**Step 3** Click Save Configuration and select Restart the affected services or instances. Click OK to restart the service for the Kafka configuration to take effect.
Step 4 Modify `fetch.message.max.bytes` in the Consumer service application to ensure that the value of `fetch.message.max.bytes` is greater than or equal to that of `message.max.bytes`.

-----End

9.16 Kafka Topics Cannot Be Deleted

Symptom

When running the following command on the Kafka client to delete topics, it is found that the topics cannot be deleted.

```
 kafka-topics.sh --delete --topic test --zookeeper 192.168.234.231:2181/kafka
```

Possible Causes

- The command for connecting the client to ZooKeeper is incorrect.
- Kafka is abnormal and some Kafka nodes are stopped.
- Perform the following operations when Kafka server configurations cannot be deleted.
- Perform the following operations when Kafka configurations are automatically created and the Producer is not stopped.

Cause Analysis

1. After the client command is run, the "ZkTimeoutException" exception is reported.

```
2016-03-09 10:41:45,773 WARN Can not get the principle name from server 192.168.234.231
(org.apache.zookeeper.ClientCnxn)
Exception in thread "main" org.I0Itec.zkclient.exception.ZkTimeoutException: Unable to connect to
zookeeper server within timeout: 30000
at org.I0Itec.zkclient.ZkClient.connect(ZkClient.java:880)
at org.I0Itec.zkclient.ZkClient.<init>(ZkClient.java:98)
at org.I0Itec.zkclient.ZkClient.<init>(ZkClient.java:84)
at kafka.admin.TopicCommand$.main(TopicCommand.scala:51)
at kafka.admin.TopicCommand.main(TopicCommand.scala)
```

For details about the solution, see Step 1.

2. Run the following query command on the client:

```
 kafka-topics.sh --list --zookeeper 192.168.0.122:2181/kafka
```

On MRS Manager, check the service status of Kafka Broker instances.

Run the `cd /var/log/Bigdata/kafka/broker` command to go to the log directory of node `RunningAsController`. Locate `ineligible for deletion: test` in the `controller.log` file.

```
2016-03-09 11:11:26,228 INFO [main] [Controller 1]: List of topics to be deleted: |
kafka.controller.KafkaController (Logging.scala:68)
```
3. On MRS Manager, view the `delete.topic.enable` status of Broker.

For details about the solution, see Step 2.

4. Run the following query command on the client:
   ```
   kafka-topics.sh --describe --topic test --zookeeper 192.168.0.122:2181/kafka
   ```

Go to the log directory of node `RunningAsController`. Locate **marked ineligible for deletion** in the `controller.log` file.

5. On FusionInsight Manager, query the Broker status.

It can be seen that a Broker is in the Stopped state. In this case, delete the topic and ensure that Brokers where partitions of the topic reside must be in the Good state.

For details about the solution, see Step 3.

6. Go to the log directory of node `RunningAsController`. Locate **Deletion successfully** in the `controller.log` file. If **New topics:[Set(test)]** is displayed again, it indicates that the topic is created again.

7. Use MRS Manager to query the topic creation configuration of Broker.
It is confirmed that the application that performs operations on the topic is not stopped.
For details about the solution, see Step 4.

Solution

**Step 1** Perform the following operations when connection to ZooKeeper fails.

When the connection between the Kafka client and ZooKeeper times out, run the ping command to check whether the Kafka client can connect to ZooKeeper.
Check the network connection between the client and ZooKeeper.
If the network connection fails, check the ZooKeeper service information on MRS Manager.

![Figure 9-26 ZooKeeper service information](image)

If ZooKeeper is improperly configured, change the ZooKeeper IP address in the client command.

**Step 2** Perform the following operations when Kafka server configurations cannot be deleted.

On MRS Manager, change the value of `delete.topic.enable` to `true`. Save the configurations and restart the service.

![Figure 9-27 Modifying delete.topic.enable.](image)
The client query command does not contain **Topic:test**.

```bash
kafka-topics.sh --list --zookeeper 192.168.0.122:24002/kafka
```

Go to the log directory of node **RunningAsController**. Locate **Deletion of topic test successfully** in the **controller.log** file.

```
2016-03-10 10:39:40,665 | INFO | [delete-topics-thread-3] | [Partition state machine on Controller 3]: Invoking state change to OfflinePartition for partitions [test,2],[test,15],[test,6],[test,16],[test,12],[test,7],[test,10],[test,13],[test,9],[test,19],[test,3],[test,1],[test,0],[test,17],[test,8],[test,4],[test,11],[test,14],[test,18] | kafka.controller.PartitionStateMachine (Logging.scala:68)
2016-03-10 10:39:40,668 | INFO | [delete-topics-thread-3] | [Partition state machine on Controller 3]: Invoking state change to NonExistentPartition for partitions [test,2],[test,15],[test,6],[test,16],[test,12],[test,7],[test,10],[test,13],[test,9],[test,19],[test,3],[test,1],[test,0],[test,17],[test,8],[test,4],[test,11],[test,14],[test,18] | kafka.controller.PartitionStateMachine (Logging.scala:68)
2016-03-10 10:39:40,977 | INFO | [delete-topics-thread-3] | [delete-topics-thread-3], Deletion of topic test successfully completed | kafka.controller.TopicDeletionManager$DeleteTopicsThread (Logging.scala:68)
```

**Step 3** Some Kafka nodes are stopped or faulty.

Start the stopped Broker instances.

The client query command does not contain **Topic:test**.

```bash
kafka-topics.sh --list --zookeeper 192.168.0.122:24002/kafka
```

Go to the log directory of node **RunningAsController**. Locate **Deletion of topic test successfully** in the **controller.log** file.

```
2016-03-10 11:17:56,463 | INFO | [delete-topics-thread-3] | [Partition state machine on Controller 3]: Invoking state change to NonExistentPartition for partitions [test,4],[test,1],[test,8],[test,2],[test,5],[test,9],[test,7],[test,6],[test,0],[test,3] | kafka.controller.PartitionStateMachine (Logging.scala:68)
2016-03-10 11:17:56,726 | INFO | [delete-topics-thread-3] | [delete-topics-thread-3], Deletion of topic test successfully completed | kafka.controller.TopicDeletionManager$DeleteTopicsThread (Logging.scala:68)
```

**Step 4** Perform the following operations when Kafka configurations are automatically created and the Producer is not stopped.

Stop related applications, change the value of **auto.create.topics.enable** to **false** on MRS Manager, save the configuration, and restart the service.

**Figure 9-28** Modifying auto.create.topics.enable

![Modifying auto.create.topics.enable](image)

**Step 5** Perform the delete operation again.

---- End
9.17 Error "AdminOperationException" Is Displayed When a Kafka Topic Is Deleted

**Symptom**

When running the following command on the Kafka client to set the ACL for a topic, it is found that the ACL cannot be set.

```
kafka-topics.sh --delete --topic test4 --zookeeper 10.5.144.2:2181/kafka
```

The error message "ERROR kafka.admin.AdminOperationException: Error while deleting topic test4" is displayed.

See the following:

```
Error while executing topic command : Error while deleting topic test4
[2017-01-25 14:00:20,750] ERROR kafka.admin/AdminOperationException: Error while deleting topic test4
```

**Possible Causes**

The user does not belong to the `kafkaadmin` group. Kafka provides a secure access interface. Only users in the `kafkaadmin` group can delete topics.

**Cause Analysis**

1. After the client command is run, the "AdminOperationException" exception is reported.

2. Run the client command `klist` to query the current authenticated user.

   ```
   [root@10-10-144-2 client]# klist
   Ticket cache: FILE:/tmp/krb5cc_0:
   Default principal: test@HADOOP.COM
   Valid starting     Expires            Service principal
   01/25/17 11:06:48  01/26/17 11:06:45  krbtgt/HADOOP.COM@HADOOP.COM
   
   In the preceding example, the authenticated user is `test`.
   ```

3. Query user group information by command ID.

   ```
   [root@10-10-144-2 client]# id test
   uid=20032(test) gid=10001(hadoop) groups=10001(hadoop),9998(ficommon),10003(kafka)
   ```

**Solution**

**Step 1** Log in to MRS Manager.

**Step 2** Choose **System > Manage User**.

**Step 3** In the **Operation** column of the user, click **Modify**.

**Step 4** Add the user to the `kafkaadmin` group.
Step 5 Run the `id` command to query the user group information.

```
[root@10-10-144-2 client]# id test
uid=20032(test) gid=10001(hadoop) groups=10001(hadoop),9998(ficommon),10002(kafkaadmin),10003(kafka)
```

----End

### 9.18 When a Kafka Topic Fails to Be Created, "NoAuthException" Is Displayed

#### Symptom

When running the following command on the Kafka client to create topics, it is found that the topics cannot be created.

```
kafka-topics.sh --create --zookeeper 192.168.234.231:2181/kafka --replication-factor 1 --partitions 2 --topic test
```

Error messages "NoAuthException" and "KeeperErrorCode = NoAuth for /config/topics" are displayed.

See the following:

```
Error while executing topic command org.apache.zookeeper.KeeperException$NoAuthException: KeeperErrorCode = NoAuth for /config/topics
```

#### Possible Causes

The user does not belong to the `kafkaadmin` group. Kafka provides a secure access interface. Only users in the `kafkaadmin` group can delete topics.
Cause Analysis

1. After the client command is run, the "NoAuthException" exception is reported.
   Error while executing topic command org.apache.zookeeper.KeeperException$NoAuthException: KeeperErrorCode = NoAuth for /config/topics
   org.I0Itec.zkclient.exception.ZkException: org.apache.zookeeper.KeeperException$NoAuthException: KeeperErrorCode = NoAuth for /config/topics
   at org.I0Itec.zkclient.exception.ZkException.create(ZkException.java:68)
   at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:685)
   at org.I0Itec.zkclient.ZkClient.create(ZkClient.java:304)
   at org.I0Itec.zkclient.ZkClient.createPersistent(ZkClient.java:213)
   at kafka.utils.ZkUtils$createParentPath(ZkUtils.scala:215)
   at kafka.utils.ZkUtils.updatePersistentPath(ZkUtils.scala:338)
   at kafka.admin.AdminUtils$writeTopicConfig(AdminUtils.scala:247)

2. Run the client command klist to query the current authenticated user.
   [root@10-10-144-2 client]# klist
   Ticket cache: FILE:/tmp/krb5cc_0
   Default principal: test@HADOOP.COM
   Valid starting Expires Service principal
   01/25/17 11:06:48  01/26/17 11:06:45  krbtgt/HADOOP.COM@HADOOP.COM
   In the preceding example, the authenticated user is test.

3. Run the id command to query the user group information.
   [root@10-10-144-2 client]# id test
   uid=20032(test) gid=10001(hadoop) groups=10001(hadoop),9998(ficommon),10003(kafka)

Solution

Step 1 Log in to MRS Manager.

Step 2 Choose System > Manage User.

Step 3 In the Operation column of the user, click Modify.

Step 4 Add the user to the kafkaadmin group.

Figure 9-30 Adding the user to the kafkaadmin group

Step 5 Run the id command to query the user group information.
9.19 Failed to Set an ACL for a Kafka Topic, and "NoAuthException" Is Displayed

Symptom

When running the following command on the Kafka client to set the ACL for a topic, it is found that the topic ACL cannot be set.

```
kafka-acls.sh --authorizer-properties zookeeper.connect=10.5.144.2:2181/kafka  --topic topic_acl --producer --add --allow-principal User:test_acl
```

The error message "NoAuthException: KeeperErrorCode = NoAuth for /kafka-acl-changes/acl_changes_0000000002" is displayed.

See the following:

```
Error while executing ACL command: org.apache.zookeeper.KeeperException$NoAuthException: KeeperErrorCode = NoAuth for /kafka-acl-changes/acl_changes_0000000002
```

```
at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:995)
at org.I0Itec.zkclient.ZkClient.delete(ZkClient.java:1038)
at kafka.utils.ZkUtils.deletePath(ZkUtils.scala:499)
at kafka.common.ZkNodeChangeNotificationListener$$anonfun$purgeObsoleteNotifications$1.apply(ZkNodeChangeNotificationListener.scala:118)
at kafka.common.ZkNodeChangeNotificationListener$$anonfun$purgeObsoleteNotifications$1.apply(ZkNodeChangeNotificationListener.scala:112)
at scala.collection.mutable.ResizableArray$class.foreach(ResizableArray.scala:59)
at scala.collection.mutable.ArrayBuffer.foreach(ArrayBuffer.scala:47)
at kafka.common.ZkNodeChangeNotificationListener.purgeObsoleteNotifications(ZkNodeChangeNotificationListener.scala:112)
at kafka.common.ZkNodeChangeNotificationListener.kafka$common$ZkNodeChangeNotificationListener$$processNotifications(ZkNodeChangeNotificationListener.scala:97)
at kafka.common.ZkNodeChangeNotificationListener.kafka$common$ZkNodeChangeNotificationListener.init(ZkNodeChangeNotificationListener.scala:77)
at kafka.security.auth.SimpleAclAuthorizer.configure(SimpleAclAuthorizer.scala:136)
at kafka.admin.AclCommand$.withAuthorizer(AclCommand.scala:136)
at kafka.admin.AclCommand$.addAcl(AclCommand.scala:48)
at kafka.admin.AclCommand$.main(AclCommand.scala)
Caused by: org.apache.zookeeper.KeeperException$NoAuthException: KeeperErrorCode = NoAuth for /kafka-acl-changes/acl_changes_0000000002
```

```
at org.apache.zookeeper.KeeperException.create(KeeperException.java:117)
at org.apache.zookeeper.KeeperException.create(KeeperException.java:51)
at org.apache.zookeeper.ZooKeeper.delete(ZooKeeper.java:1416)
at org.I0Itec.zkclient.ZkConnection.delete(ZkConnection.java:104)
at org.I0Itec.zkclient.ZkClient$1.call(ZkClient.java:1042)
at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:985)
```
Possible Causes

The user does not belong to the kafkaadmin group. Kafka provides a secure access interface. Only users in the kafkaadmin group can perform the setting operation.

Cause Analysis

1. After the client command is run, the "NoAuthException" exception is reported.
2. Run the client command klist to query the current authenticated user.
   
   ```
   [root@10-10-144-2 client]# klist
   Ticket cache: FILE:/tmp/krb5cc_0
   Default principal: test@HADOOP.COM
   
   Valid starting     Expires            Service principal
   01/25/17 11:06:48  01/26/17 11:06:45  krbtgt/HADOOP.COM@HADOOP.COM
   
   In the preceding example, the authenticated user is test.
   
   3. Run the id command to query the user group information.
   ```

Solution

**Step 1** Log in to MRS Manager.

**Step 2** Choose System > Manage User.

**Step 3** In the Operation column of the user, click Modify.

**Step 4** Add the user to the kafkaadmin group.

**Figure 9-31 Modifying a user group**

**Step 5** Run the id command to query the user group information.

```
[root@host1 client]# id test
uid=20032(test) gid=10001(hadoop) groups=10001(hadoop),9998(ficommon),10003(kafka)

----End
9.20 When a Kafka Topic Fails to Be Created, "NoNode for /brokers/ids" Is Displayed

**Symptom**

When running the following command on the Kafka client to create topics, it is found that the topics cannot be created.

```
kafka-topics.sh --create  --replication-factor 1 --partitions 2 --topic test --zookeeper 192.168.234.231:2181
```

The error message "NoNodeException: KeeperErrorCode = NoNode for /brokers/ids" is displayed.

See the following:

```
Error while executing topic command : org.apache.zookeeper.KeeperException$NoNodeException: KeeperErrorCode = NoNode for /brokers/ids
  at org.I0Itec.zkclient.KafkaException.create(KafkaException.java:47)
  at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:995)
  at org.I0Itec.zkclient.ZkClient.getZkClient(ZkClient.java:671)
  at kafka.utils.ZkUtils.getZkUtils(ZkUtils scala:541)
  at kafka.utils.ZkUtils.getSortedBrokerList(ZkUtils.scala:176)
  at kafka.admin.TopicCommand$.createTopic(TopicCommand.scala:105)
  at kafka.admin.TopicCommand$.main(TopicCommand.scala:60)
Caused by: org.apache.zookeeper.KeeperException$NoNodeException: KeeperErrorCode = NoNode for /brokers/ids
  at org.apache.zookeeper.KeeperException.create(KeeperException.java:115)
  at org.apache.zookeeper.ZooKeeper.getChildren(ZooKeeper.java:2256)
  at org.apache.zookeeper.ZooKeeper.getChildren(ZooKeeper.java:2284)
  at org.I0Itec.zkclient.ZkClient$4.call(ZkClient.java:114)
  at org.I0Itec.zkclient.ZkClient$4.call(ZkClient.java:678)
  at org.I0Itec.zkclient.ZkClient$4.call(ZkClient.java:675)
  at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:985)
  ... 8 more
(kafka.admin.TopicCommand$)
```

**Possible Causes**

- The Kafka service is not running.
- The ZooKeeper address parameter in the client command is incorrectly configured.

**Cause Analysis**

1. After the client command is run, the "NoNodeException" exception is reported.

```
Error while executing topic command : org.apache.zookeeper.KeeperException$NoNodeException: KeeperErrorCode = NoNode for /brokers/ids
[2017-09-17 16:35:28,520] ERROR org.I0Itec.zkclient.KafkaException$KafkaException: KeeperException$KafkaException: KeeperErrorCode = NoNode for /brokers/ids
  at org.I0Itec.zkclient.KafkaException.create(KafkaException.java:47)
  at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:995)
  at org.I0Itec.zkclient.ZkClient$4.call(ZkClient.java:114)
  at org.I0Itec.zkclient.ZkClient$4.call(ZkClient.java:678)
  at org.I0Itec.zkclient.ZkClient$4.call(ZkClient.java:675)
  at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:985)
  ... 8 more
```

 организация Huawei Technologies Co., Ltd.
2. Check whether the Kafka service is in the normal state on MRS Manager.

**Figure 9-32 Kafka status**

3. Check whether the ZooKeeper address in the client command is correct. Check the Kafka information stored in ZooKeeper. The path (Znode) should be suffixed with `/kafka`. It is found that `/kafka` is missing in the configuration.

```
[root@10-10-144-2 client]# kafka-topics.sh --create --replication-factor 1 --partitions 2 --topic test --zookeeper 192.168.234.231:2181
```

**Solution**

**Step 1** Ensure that the Kafka service is normal.

**Step 2** Add `/kafka` to the ZooKeeper address in the command.

```
[root@10-10-144-2 client]# kafka-topics.sh --create --replication-factor 1 --partitions 2 --topic test --zookeeper 192.168.234.231:24002/kafka
```

----End
9.21 When a Kafka Topic Fails to Be Created, "replication factor larger than available brokers" Is Displayed

Symptom

When running the following command on the Kafka client to create topics, it is found that the topics cannot be created.

```bash
kafka-topics.sh --create --replication-factor 2 --partitions 2 --topic test --zookeeper 192.168.234.231:2181
```

The error message "replication factor larger than available brokers" is displayed.

See the following:

Error while executing topic command : replication factor: 2 larger than available brokers: 0
[2017-09-17 16:44:12,396] ERROR kafka.admin.AdminOperationException: replication factor: 2 larger than available brokers: 0
at kafka.admin.AdminUtils$.assignReplicasToBrokers(AdminUtils.scala:117)
at kafka.admin.AdminUtils$.createTopic(AdminUtils.scala:403)
at kafka.admin.TopicCommand$.createTopic(TopicCommand.scala:110)
at kafka.admin.TopicCommand$.main(TopicCommand.scala:61)
at kafka.admin.TopicCommand.main(TopicCommand.scala)
(kafka.admin.TopicCommand$)

Possible Causes

- The Kafka service is not running.
- The available Broker of the Kafka service is smaller than the configured `replication-factor`.
- The ZooKeeper address parameter in the client command is incorrectly configured.

Cause Analysis

1. After the client command is run, "replication factor larger than available brokers" is reported.

   Error while executing topic command : replication factor: 2 larger than available brokers: 0
   [2017-09-17 16:44:12,396] ERROR kafka.admin.AdminOperationException: replication factor: 2 larger than available brokers: 0
   at kafka.admin.AdminUtils$.assignReplicasToBrokers(AdminUtils.scala:117)
at kafka.admin.AdminUtils$.createTopic(AdminUtils.scala:403)
at kafka.admin.TopicCommand$.createTopic(TopicCommand.scala:110)
at kafka.admin.TopicCommand$.main(TopicCommand.scala:61)
at kafka.admin.TopicCommand.main(TopicCommand.scala)
   (kafka.admin.TopicCommand$)

2. Check whether the Kafka service is in the normal state on MRS Manager and whether the current available Broker is smaller than the configured `replication-factor`. 
3. Check whether the ZooKeeper address in the client command is correct. Check the Kafka information stored in ZooKeeper. The path (Znode) should be suffixed with `/kafka`. It is found that `/kafka` is missing in the configuration.

```bash
[root@10-10-144-2 client]#
kafka-topics.sh --create  --replication-factor 2 --partitions 2 --topic test --zookeeper 192.168.234.231:2181
```

**Solution**

**Step 1** Ensure that the Kafka service is in the normal state and the available Broker is not less than the configured `replication-factor`. 
Step 2 Add /kafka to the ZooKeeper address in the command.

```
[root@10-10-144-2 client]# kafka-topics.sh --create --replication-factor 1 --partitions 2 --topic test --zookeeper 192.168.234.231:2181/kafka
```

-----End

### 9.22 Consumer Repeatedly Consumes Data

#### Symptom

When the data volume is large, rebalance occurs frequently, causing repeated consumption. The key logs are as follows:

2018-05-12 10:58:42,561 | INFO | [kafka-request-handler-3] | [GroupCoordinator 2]: Preparing to restabilize group DemoConsumer with old generation 118 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
2018-05-12 10:58:43,560 | INFO | [kafka-request-handler-7] | [GroupCoordinator 2]: Assignment received from leader for group DemoConsumer for generation 119 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
2018-05-12 10:59:13,562 | INFO | [kafka-request-handler-3] | [GroupCoordinator 2]: Preparing to restabilize group DemoConsumer with old generation 119 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
2018-05-12 10:59:13,790 | INFO | [kafka-request-handler-3] | [GroupCoordinator 2]: Stabilized group DemoConsumer generation 120 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
2018-05-12 10:59:13,791 | INFO | [kafka-request-handler-0] | [GroupCoordinator 2]: Assignment received from leader for group DemoConsumer for generation 120 | kafka.coordinator.GroupCoordinator (Logging.scala:68)
2018-05-12 10:59:52,456 | INFO | [group-metadata-manager-0] | [Group Metadata Manager on Broker 2]: Removed 0 expired offsets in 0 milliseconds. | kafka.coordinator.GroupMetadataManager (Logging.scala:68)
2018-05-12 11:00:49,772 | INFO | [kafka-scheduler-6] | Deleting segment 0 from log '__consumer_offsets-17' | kafka.log.Log (Logging.scala:68)
In the logs, "Preparing to restabilize group DemoConsumer with old generation" indicates that rebalance occurs.

### Possible Causes

The parameter settings are improper.

### Cause Analysis

Cause: Due to improper parameter settings, the data processing time is too long when the data volume is large. Balance frequently occurs, and the offset cannot be submitted normally. As a result, the data is repeatedly consumed.

Principle: The offset is submitted only after the poll data is processed. If the processing duration after the poll data is processed exceeds the duration specified by `session.timeout.ms`, the rebalance occurs. As a result, the consumption fails and the offset of the consumed data cannot be submitted. Therefore, the data is consumed at the old offset next time. As a result, the data is repeatedly consumed.

### Solution

Adjust the following service parameters on MRS Manager:

- `request.timeout.ms=100000`
- `session.timeout.ms=90000`
- `max.poll.records=50`
- `heartbeat.interval.ms=3000`

Among the preceding parameters:

The value of `request.timeout.ms` is 10s greater than that of `session.timeout.ms`.

The value of `session.timeout.ms` must be within the values of `group.min.session.timeout.ms` and `group.max.session.timeout.ms` on the server.
Set the parameters as required. The `max.poll.records` parameter specifies the number of records for each poll. The purpose is to ensure that the processing time of poll data does not exceed the value of `session.timeout.ms`.

**Related Information**
- The post-poll data processing must be efficient and do not block the next poll.
- The poll method and data processing suggestion are processed asynchronously.

### 9.23 Partitions Are Unevenly Distributed on Nodes During the Creation of Kafka Topics

**Symptom**

When a user creates a topic using the Kafka client command, partitions are unevenly distributed.

**Possible Causes**

Racks are set for Kafka nodes and each rack houses different number of nodes.

**Cause Analysis**

1. Check the Broker instance node information about the Kafka service on MRS Manager. It is found that the number of nodes varies with racks.
2. It is found that a large number of partitions are allocated to the Broker instance on the rack node.

**Solution**

**Step 1**
Ensure that each rack houses the same number of Broker instances.

**Step 2**
When creating a topic, disable the rack awareness feature and add the parameter `--disable-rack-aware`.

The following provides an example:

```
```

----End

### 9.24 Leader for the Created Kafka Topic Partition Is Displayed as none

**Symptom**

When a user creates a topic using the Kafka client command, the leader for the created topic partition is displayed as `none`.

---
Possible Causes

- The Kafka service is not running.
- The user group information cannot be found.

Cause Analysis

1. On the MRS Manager page, choose Services > Kafka. The status of Kafka is Good and the monitoring indicators show no errors.

2. On MRS Manager, choose Services > Kafka and obtain the Controller node information from Kafka Summary.

Figure 9-35 Kafka service information

3. Log in to the node where the Controller resides and run the cd /var/log/Bigdata/kafka/broker command to go to the node log directory. In state-change.log, the message "NoAuthException" is displayed, indicating that the ZooKeeper permission is abnormal.
org.I0Itec.zkclient.exception.ZkException: org.apache.zookeeper.KeeperException$NoAuthException: KeeperErrorCode = NoAuth for /brokers/topics/partitions
at org.I0Itec.zkclient.exception.ZkException.create(ZkException.java:68)
at org.I0Itec.zkclient.ZkClient.retryUntilConnected(ZkClient.java:1000)
at org.I0Itec.zkclient.ZkClient.create(ZkClient.java:527)
at org.I0Itec.zkclient.ZkClient.createPersistent(ZkClient.java:293)

4. Check on ZooKeeper audit logs recorded in the specified period also indicates that the permission is abnormal.

2018-05-31 09:20:42,421 | ERROR | CommitProcWorkThread-1 | session=0xc3000007015d5a18
user=10.6.92.39,kafka/hadoop.hadoop.com@HADOOP.COM,kafka/
hadoop.hadoop.com@HADOOP.COM ip=10.6.92.39 operation=create znode
target=ZooKeeperServer znode=/kafka/brokers/topics/test/partitions/0/state result=failure
2018-05-31 09:20:42,423 | ERROR | CommitProcWorkThread-1 | session=0xc3000007015d5a18
user=10.6.92.39,kafka/hadoop.hadoop.com@HADOOP.COM,kafka/
hadoop.hadoop.com@HADOOP.COM ip=10.6.92.39 operation=create znode
target=ZooKeeperServer znode=/kafka/brokers/topics/partitions/0 result=failure
2018-05-31 09:20:42,435 | ERROR | CommitProcWorkThread-1 | session=0xc3000007015d5a18
user=10.6.92.39,kafka/hadoop.hadoop.com@HADOOP.COM,kafka/
hadoop.hadoop.com@HADOOP.COM ip=10.6.92.39 operation=create znode
target=ZooKeeperServer znode=/kafka/brokers/topics/partitions/0 state result=failure
2018-05-31 09:20:42,439 | ERROR | CommitProcWorkThread-1 | session=0xc3000007015d5a18
user=10.6.92.39,kafka/hadoop.hadoop.com@HADOOP.COM,kafka/
hadoop.hadoop.com@HADOOP.COM ip=10.6.92.39 operation=create znode
target=ZooKeeperServer znode=/kafka/brokers/topics/partitions/0 result=failure
2018-05-31 09:20:42,441 | ERROR | CommitProcWorkThread-1 | session=0xc3000007015d5a18
user=10.6.92.39,kafka/hadoop.hadoop.com@HADOOP.COM,kafka/
hadoop.hadoop.com@HADOOP.COM ip=10.6.92.39 operation=create znode
target=ZooKeeperServer znode=/kafka/brokers/topics/partitions/0 state result=failure

5. Run the `id -Gn kafka` command on each ZooKeeper instance node. It is found that user group information cannot be queried on a node.

```
[root @bdpsit3ap03 ~]# id -Gn kafka
id: kafka: No such user
[root @bdpsit3ap03 ~]#
```

6. In an MRS cluster, user management is provided by the LDAP service and depends on the SSSD (Red Hat) and NSCD (SUSE) services of OSs. The process from creating a user to synchronizing the user to the SSSD service takes some time. If the user does not take effect or the SSSD version has bugs, the user may be invalid on the ZooKeeper node in some cases, which causes topic creation failures.

**Solution**

**Step 1** Restart the SSD/NSCD service.
- **RedHat**
  
  service sssd restart
- **SUSE**
  
  service nscd restart

**Step 2** After restarting related services, run the `id username` command on the active ResourceManager node to check whether the user information is valid.

----End
9.25 Safety Instructions on Using Kafka

Brief Introduction to API for Kafka

- New Producer API
  Indicates the API defined in org.apache.kafka.clients.producer.KafkaProducer. When kafka-console-producer.sh is used, the API is used by default.

- Old Producer API
  Indicates the API defined in kafka.producer.Producer. When kafka-console-producer.sh is used, the API is invoked to add --old-producer.

- New Consumer API
  Indicates the API defined in org.apache.kafka.clients.consumer.KafkaConsumer. When kafka-console-consumer.sh is used, the API is invoked to add --new-consumer.

- Old Consumer API
  Indicates the API defined in kafka.consumer.ConsumerConnector. When kafka-console-consumer.sh is used, the API is used by default.

**NOTE**

New Producer API and new Consumer API are called new API in general in the document.

Protocol Description for Accessing Kafka

The protocols used to access Kafka are as follows: PLAINTEXT, SSL, SASL_PLAINTEXT, and SASL_SSL.

When Kafka service is started, the listeners using the PLAINTEXT and SASL_PLAINTEXT protocols are started. You can set ssl.mode.enable to true in Kafka service configuration to start listeners using SSL and SASL_SSL protocols.

The following table describes the four protocols:

<table>
<thead>
<tr>
<th>Protocol Type</th>
<th>Description</th>
<th>Supported API</th>
<th>Default Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLAINTEXT</td>
<td>Supports plaintext access without authentication.</td>
<td>New and old APIs</td>
<td>21005</td>
</tr>
<tr>
<td>SASL_PLAINTEXT</td>
<td>Supports plaintext access with Kerberos authentication.</td>
<td>New API</td>
<td>21007</td>
</tr>
<tr>
<td>SSL</td>
<td>Supports SSL-encrypted access without authentication.</td>
<td>New API</td>
<td>21008</td>
</tr>
<tr>
<td>SASL_SSL</td>
<td>Supports SSL-encrypted access with Kerberos authentication.</td>
<td>New API</td>
<td>21009</td>
</tr>
</tbody>
</table>
ACL Settings for Topic

Kafka supports secure access. Therefore, users can set the ACL for topics to control that different users access different topics. To view and set the permission information about a topic, run the kafka-acls.sh script on the Linux client.

- Scenarios
  Assign Kafka users with specific permissions for related topics based on service requirements.

The following table describes default Kafka user groups.

<table>
<thead>
<tr>
<th>User Group</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>kafkaadmin</td>
<td>Kafka administrator group. Users added to this group have the permissions to create, delete, authorize, as well as read from and write data to all topics.</td>
</tr>
<tr>
<td>kafkasuperuser</td>
<td>Indicates a super Kafka user group. Users added to this group have permissions to read data from and write data to all topics.</td>
</tr>
<tr>
<td>kafka</td>
<td>Kafka common user group. If users in this group want to read data from and write data to a specific topic, the users in the kafkaadmin group must grant permissions to users in this group.</td>
</tr>
</tbody>
</table>

- Prerequisites
  a. The system administrator has understood service requirements and prepared a Kafka administrator (belonging to the kafkaadmin group).
  b. The Kafka client has been installed.

- Procedure
  a. Use PuTTY to log in to the node where the Kafka client is installed as the client installation user.
  b. Switch to the Kafka client installation directory, for example, /opt/kafkaclient.
     `cd /opt/kafkaclient`
  c. Run the following command to configure environment variables:
     `source bigdata_env`
  d. Run the following command to perform user authentication (Skip this step for a common cluster.)
     `kinit Component user`
  e. Run the following command to switch to the Kafka client installation directory:
     `cd Kafka/kafka/bin`
f. The following describes the commands commonly used for user authorization when `kafka-acl.sh` is used:

- View the permission control list of a topic:
  
  ```
  ./kafka-acls.sh --authorizer-properties
  zookeeper.connect=<ZooKeeper cluster service IP address:2181/
  kafka > --list --topic <Topic name>
  ```

- Add the Producer permission for a user:
  
  ```
  ./kafka-acls.sh --authorizer-properties
  zookeeper.connect=<ZooKeeper cluster service IP address:2181/
  kafka > --add --allow-principal User:<User name> --producer --topic <Topic name>
  ```

- Remove the Producer permission from a user:
  
  ```
  ./kafka-acls.sh --authorizer-properties
  zookeeper.connect=<ZooKeeper cluster service IP address:2181/
  kafka > --remove --allow-principal User:<User name> --producer --topic <Topic name>
  ```

- Add the Consumer permission for a user:
  
  ```
  ./kafka-acls.sh --authorizer-properties
  zookeeper.connect=<ZooKeeper cluster service IP address:2181/
  kafka > --add --allow-principal User:<User name> --consumer --topic <Topic name> --group<Consumer group name>
  ```

- Remove the Consumer permission from a user:
  
  ```
  ./kafka-acls.sh --authorizer-properties
  zookeeper.connect=<ZooKeeper cluster service IP address:2181/
  kafka > --remove --allow-principal User:<User name> --consumer --topic <Topic name> --group<Consumer group name>
  ```

**NOTE**

In MRS 1.6.3 or earlier, the default ZooKeeper port number is 24002 regardless of whether Kerberos authentication is enabled for the cluster. In MRS 1.6.3 or later, the default ZooKeeper port number is 2181 regardless of whether Kerberos authentication is enabled for the cluster.

---

**Use of New and Old Kafka APIs in Different Scenarios**

- **Scenario 1: accessing the topic with an ACL**

<table>
<thead>
<tr>
<th>Used API</th>
<th>User Group</th>
<th>Client Parameter</th>
<th>Server Parameter</th>
<th>Access Port</th>
</tr>
</thead>
</table>
| New API   | Users need to meet one of the following conditions:  
  - In the administrator | security.protocol=SASL_PLAINTEXT  
  sasl.kerberos.service.name = kafka | - | sasl.port (The default number is 21007.) |
### Scenario 1: accessing the topic with an ACL

<table>
<thead>
<tr>
<th>Used API</th>
<th>User Group</th>
<th>Client Parameter</th>
<th>Server Parameter</th>
<th>Access Port</th>
</tr>
</thead>
</table>
| New API  | Users need to meet one of the following conditions:  
- In the administrator group  
- In the *kafkaadmin* group  
- In the *kafkasuperuser* group  
- In the *kafka* group and be authorized | security.protocol=SASL_SSL  
sasl.kerberos.service.name=kafka | Set ssl.mode.enable to true. | sasl-ssl.port (The default port number is 21009.) |

Old API: Access is not supported.

### Scenario 2: accessing the topic without an ACL

<table>
<thead>
<tr>
<th>Used API</th>
<th>User Group</th>
<th>Client Parameter</th>
<th>Server Parameter</th>
<th>Access Port</th>
</tr>
</thead>
</table>
| New API  | Users need to meet one of the following conditions:  
- In the administrator group  
- In the *kafkaadmin* group  
- In the *kafkasuperuser* group  
- In the *kafka* group and be authorized | security.protocol=SASL_PLAINTEXT  
sasl.kerberos.service.name=kafka | - | sasl.port (The default number is 21007.) |
<table>
<thead>
<tr>
<th>Used API</th>
<th>User Group</th>
<th>Client Parameter</th>
<th>Server Parameter</th>
<th>Access Port</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Users are in the <strong>kafka</strong> group.</td>
<td></td>
<td>Set <code>allow.everyone.if.no.acl.found</code> to true.</td>
<td><code>sasl.port</code> (The default number is 21007.)</td>
</tr>
<tr>
<td></td>
<td>Users need to meet one of the following conditions:</td>
<td><code>security.protocol=SASL_SSL,sasl.kerberos.service.name=kafka</code></td>
<td>Set <code>ssl-enable</code> to true.</td>
<td><code>sasl-ssl.port</code> (The default port number is 21009.)</td>
</tr>
<tr>
<td></td>
<td>● In the administrator group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>● In the <strong>kafkaadmin</strong> group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>● In the <strong>kafkasupervisor</strong> group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Users are in the <strong>kafka</strong> group.</td>
<td></td>
<td>Set <code>allow.everyone.if.no.acl.found</code> to true.</td>
<td><code>sasl-ssl.port</code> (The default port number is 21009.)</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td><code>security.protocol=PLAIN TEXT</code></td>
<td>Set <code>allow.everyone.if.no.acl.found</code> to true.</td>
<td><code>port</code> (The default number is 21005.)</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td><code>security.protocol=SSL</code></td>
<td>Set <code>allow.everyone.if.no.acl.found</code> to true.</td>
<td><code>ssl.port</code> (The default number is 21008.)</td>
</tr>
<tr>
<td>Old Producer</td>
<td>-</td>
<td>-</td>
<td>Set <code>allow.everyone.if.no.acl.found</code> to true.</td>
<td><code>port</code> (The default number is 21005.)</td>
</tr>
</tbody>
</table>
9.26 Obtaining Kafka Consumer Offset Information

Symptom

How do I obtain Kafka Consumer offset information when using Kafka Consumer to consume data?

Kafka APIs

- New Producer API
  Indicates the API defined in `org.apache.kafka.clients.producer.KafkaProducer`. When `kafka-console-producer.sh` is used, the API is used by default.

- Old Producer API
  Indicates the API defined in `kafka.producer.Producer`. When `kafka-console-producer.sh` is used, the API is invoked to add `--old-producer`.

- New Consumer API
  Indicates the API defined in `org.apache.kafka.clients.consumer.KafkaConsumer`. When `kafka-console-consumer.sh` is used, the API is invoked to add `--new-consumer`.

- Old Consumer API
  Indicates the API defined in `kafka.consumer.ConsumerConnector`. When `kafka-console-consumer.sh` is used, the API is used by default.

**NOTE**

New Producer API and new Consumer API are called new API in general in the document.

Procedure

Old Consumer API

- Prerequisites
  a. The system administrator has understood service requirements and prepared a Kafka administrator (belonging to the kafkaadmin group).
  b. The Kafka client has been installed.
- Procedure
a. Use PuTTY to log in to the node where the Kafka client is installed as the client installation user.

b. Switch to the Kafka client installation directory, for example, /opt/kafkaclient.

cd /opt/kafkaclient

c. Run the following command to configure environment variables:

source bigdata_env

d. Run the following command to perform user authentication (skip this step in common mode):

kinit Component service user

e. Run the following command to switch to the Kafka client installation directory:

cd Kafka/kafka/bin

f. Run the following command to obtain Consumer offset metric information:

```bash
bin/kafka-consumer-groups.sh --zookeeper <zookeeper_host:port>/kafka --list
```

```bash
bin/kafka-consumer-groups.sh --zookeeper <zookeeper_host:port>/kafka --describe --group test-consumer-group
```

The following provides an example:

```bash
kafka-consumer-groups.sh --zookeeper 192.168.100.100:2181/kafka --list
```

```bash
kafka-consumer-groups.sh --zookeeper 192.168.100.100:2181/kafka --describe --group test-consumer-group
```

New Consumer API

- **Prerequisites**
  
a. The system administrator has understood service requirements and prepared a Kafka administrator (belonging to the kafkaadmin group).

b. The Kafka client has been installed.

- **Procedure**
  
a. Use PuTTY to log in to the node where the Kafka client is installed as the client installation user.

b. Switch to the Kafka client installation directory, for example, /opt/client.

cd /opt/client

c. Run the following command to configure environment variables:

source bigdata_env

d. Run the following command to perform user authentication (skip this step in common mode):

kinit Component service user

e. Run the following command to switch to the Kafka client installation directory:

cd Kafka/kafka/bin

f. Run the following command to obtain Consumer offset metric information:

```bash
kafka-consumer-groups.sh --bootstrap-server <broker_host:port> --describe --group my-group
```
The following provides an example:
```
kafka-consumer-groups.sh --bootstrap-server 192.168.100.100:9092 --describe --group my-group
```

## 9.27 Adding or Deleting Configurations for a Topic

### Symptom

Configure or modify a specific topic when using Kafka.

### Parameters that can be modified at the topic level:

- cleanup.policy
- compression.type
- delete.retention.ms
- file.delete.delay.ms
- flush.messages
- flush.ms
- index.interval.bytes
- max.message.bytes
- min.cleanable.dirty.ratio
- min.insync.replicas
- preallocate
- retention.bytes
- retention.ms
- segment.bytes
- segment.index.bytes
- segment.jitter.ms
- segment.ms
- unclean.leader.election.enable

### Procedure

- **Prerequisites**
  The Kafka client has been installed.

- **Procedure**
  a. Use PuTTY to log in to the node where the Kafka client is installed as the client installation user.
  b. Switch to the Kafka client installation directory, for example, `/opt/client`.
  c. Run the following command to configure environment variables:
     ```
     source bigdata_env
     ```
  d. Run the following command to perform user authentication (skip this step in common mode):
     ```
     kinit Component service user
     ```
  e. Run the following command to switch to the Kafka client installation directory:
     ```
     cd Kafka/kafka/bin
     ```
  f. Run the following commands to configure and delete a topic:
     ```
     kafka-topics.sh --alter --topic <topic_name> --zookeeper <zookeeper_host:port>/kafka --config <name=value>
     kafka-topics.sh --alter --topic <topic_name> --zookeeper <zookeeper_host:port>/kafka --delete-config <name>
The following provides an example:
```
kafka-topics.sh --alter --topic test1 --zookeeper 192.168.100.100:2181/kafka --config retention.ms=86400000
kafka-topics.sh --alter --topic test1 --zookeeper 192.168.100.100:2181/kafka --delete-config retention.ms
```
g. Run the following command to query topic information:
```
kafka-topics.sh --describe -topic <topic_name> --zookeeper <zookeeper_host:port>/kafka
```

9.28 Reading the Content of the __consumer_offets Internal Topic

Issue

How does Kafka save the offset of a Consumer to the __consumer_offets of internal topics?

Procedure

Step 1  Use PuTTY to log in to the node where the Kafka client is installed as the client installation user.

Step 2  Switch to the Kafka client installation directory, for example, /opt/client.
```
cd /opt/client
```

Step 3  Run the following command to configure environment variables:
```
source bigdata_env
```

Step 4  Run the following command to perform user authentication (Skip this step for a common cluster.)
```
kinit Component service user
```

Step 5  Run the following command to switch to the Kafka client installation directory:
```
cd Kafka/kafka/bin
```

Step 6  Run the following command to obtain Consumer offset metric information:
```
kafka-console-consumer.sh --topic __consumer_offsets --zookeeper <zk_host:port>/kafka --formatter "kafka.coordinator.group.GroupMetadataManager\ SOffsetsMessageFormatter" --consumer.config <property file> --from-beginning
```

Add the following content to the <property file> configuration file:
```
exclude.internal.topics = false
```

The following provides an example:
```
kafka-console-consumer.sh --topic __consumer_offsets --zookeeper 10.5.144.2:2181/kafka --formatter
```
9.29 Configuring Logs for Shell Commands on the Client

**Issue**

How do I set the log level for shell commands on the client?

**Procedure**

**Step 1** Use PuTTY to log in to the node where the Kafka client is installed as the client installation user.

**Step 2** Switch to the Kafka client installation directory, for example, `/opt/client`.

```
cd /opt/client
```

**Step 3** Run the following command to switch to the Kafka client configuration directory:

```
cd Kafka/kafka/config
```

**Step 4** Open the `tools-log4j.properties` file, change `WARN` to `INFO`, and save the file.

```
log4j.rootLogger=WARN, stderr
log4j.appenders.stderr=org.apache.log4j.ConsoleAppender
log4j.appenders.stderr.layout=org.apache.log4j.PatternLayout
log4j.appenders.stderr.layout.ConversionPattern=%d [%t] %-5p %c\n
log4j.rootLogger=INFO, stderr
log4j.appenders.stderr=org.apache.log4j.ConsoleAppender
log4j.appenders.stderr.layout=org.apache.log4j.PatternLayout
log4j.appenders.stderr.layout.ConversionPattern=%d [%t] %-5p %c\n```

**Step 5** Switch to the Kafka client installation directory, for example, `/opt/client`.

```
cd /opt/client
```

**Step 6** Run the following command to configure environment variables:

```
source bigdata_env
```

**Step 7** Run the following command to perform user authentication (Skip this step for a common cluster.)

```
kinit Component service user
```
Step 8 Run the following command to switch to the Kafka client installation directory:

```
cd Kafka/kafka/bin
```

Step 9 Run the following command to obtain the topic information. The log information can be viewed on the console.

```
kafka-topics.sh --list --zookeeper 10.5.144.2:2181/kafka
```

9.30 Obtaining Topic Distribution Information

Issue

How do I obtain topic distribution information in a Broker instance?

Preparations

- Prerequisites
  The Kafka and ZooKeeper clients have been installed.
- Procedure
  a. Use PuTTY to log in to the node where the Kafka client is installed as the client installation user.
  b. Switch to the Kafka client installation directory, for example, `/opt/client`.
  
  ```
cd /opt/client
  ```
  c. Run the following command to configure environment variables:
  
  ```
  source bigdata_env
  ```
  d. Run the following command to perform user authentication (Skip this step for a common cluster.)
  
  ```
kinit Component user
  ```
  e. Run the following command to switch to the Kafka client installation directory:
  
  ```
cd Kafka/kafka/bin
  ```
  f. Run the Kafka commands to obtain the topic assignment information and copy synchronization information, and check the return result.
kafka-topics.sh --describe --zookeeper <zk_host:port/chroot>

The following provides an example:

```
[root@mgtdat-sh-3-01-3 client]# kafka-topics.sh --describe --zookeeper 10.149.0.90:2181/kafka

Topic: topic1  PartitionCount: 2  ReplicationFactor: 2
Configs:

Topic: topic1  Partition: 1  Leader: 24  Replicas: 24, 23  Isr: 24, 23
```

In the preceding information, **Replicas** indicates the replica assignment information and **Isr** indicates the replica synchronization information.

**Solution 1**

1. Query the Broker ID mapping in ZooKeeper.
   
   ```
   sh zkcli.sh -server <zk_host:port>
   ```
   
2. Run the following command on the ZooKeeper client:
   
   ```
   ls /kafka/brokers/ids
   ```
   
   ```
   get/kafka/brokers/ids/<queried Broker ID>
   ```
   
   The following provides an example:
   ```
   [root@node-master1gAMQ kafka]# zkcli.sh -server node-master1gAMQ:2181
   Connecting to node-master1gAMQ:2181
   Welcome to ZooKeeper!
   JLine support is enabled
   
   WATCHER::
   
   WatchedEvent state:SyncConnected type:None path:null
   [zk: node-master1gAMQ:2181(CONNECTED) 0] ls /kafka/brokers/ids
   [zk: node-master1gAMQ:2181(CONNECTED) 0] ls /kafka/brokers/ids/1
   
   ```
   ```
   
   [zk: node-master1gAMQ:2181(CONNECTED) 1] get /kafka/brokers/ids/1
   
   ```
   ```
   [zk: node-master1gAMQ:2181(CONNECTED) 2]
   ```

**Solution 2**

Obtain the mapping between nodes and Broker IDs.

```
kafka-broker-info.sh --zookeeper <zk_host:port/chroot>
```

The following provides an example:

```
[root@node-master1gAMQ kafka]# bin/kafka-broker-info.sh --zookeeper 192.168.2.70:2181/kafka

Broker_ID   IP_Address

1       192.168.2.242
```

### 9.31 Kafka HA Usage Description

**Kafka High Reliability and Availability**

Kafka message transmission assurance mechanism ensures message transmission after required parameters are set to meet different performance and reliability requirements.
Kafka high availability and high performance

If HA and high performance are required, configure parameters listed in the following table.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Default Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>unclean.leader.election.enable</td>
<td>true</td>
<td>Specifies whether a replica that is not in the ISR can be selected as the leader. If this parameter is set to <strong>true</strong>, data may be lost.</td>
</tr>
<tr>
<td>auto.leader.rebalance.enable</td>
<td>true</td>
<td>Specifies whether the leader automated balancing function is used.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>If this parameter is set to <strong>true</strong>, the controller periodically balances the leader of each partition on all nodes and assigns the leader to a replica with a higher priority.</td>
</tr>
<tr>
<td>acks</td>
<td>1</td>
<td>The leader needs to check whether the message has been received and determine whether the required operation has been processed. This parameter affects message reliability and performance.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• If this parameter is set to <strong>0</strong>, the Producer does not wait for any response from the server, and the message is considered successful.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• If this parameter is set to <strong>1</strong>, when the leader of the copy verifies that data has been written into the cluster, the leader makes repose quickly without waiting until all the copies are written. In this case, if the leader is abnormal when the leader makes the confirmation but replica synchronization is not complete, data will be lost.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• If this parameter is set to <strong>-1</strong> (all), the synchronization is successful only after all synchronization copies are confirmed. If <strong>min.insync.replicas</strong> is also configured, multiple copies can be written successfully. In this case, as long as one copy remains active, the record is not lost.</td>
</tr>
<tr>
<td>min.insync.replicas</td>
<td>1</td>
<td>Specifies the minimum number of replicas to which data is written when <strong>acks</strong> is set to <strong>-1</strong> for the Producer.</td>
</tr>
</tbody>
</table>
After HA and high performance are configured, the data reliability decreases. Specifically, data may be lost if disks or nodes are faulty.

- **Kafka high reliability configuration**

  If high data reliability is required, configure parameters listed in the following table.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Recommended Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>unclean.leader.election.enable</td>
<td>false</td>
<td>Indicates whether a replica that is not in the ISR list can be elected as a leader.</td>
</tr>
<tr>
<td>acks</td>
<td>-1</td>
<td>The leader needs to check whether the message has been received and determine whether the required operation has been processed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>If this parameter is set to -1, the message is considered to be successfully received only when all replicas in the ISR list have confirmed to receive the message. The min.insync.replicas parameter must also be set to ensure that multiple copies can be written successfully. As long as one copy is active, the record is not lost.</td>
</tr>
<tr>
<td>min.insync.replicas</td>
<td>2</td>
<td>Specifies the minimum number of replicas to which data is written when acks is set to -1 for the Producer. Ensure that the value of Min.insync.replicas is equal to or less than that of replication.factor.</td>
</tr>
</tbody>
</table>

Impact of high reliability configurations:

- **Deteriorated performance**

  All copies in the ISR list are required, and the writing of the minimum number of copies has been verified successful. As a result, the delay of a single message increases and the processing capability of the client decreases. The actual performance depends on the onsite test data.

- **Reduced availability**

  A replica that is not in the ISR list cannot be elected as a leader. If the leader goes offline and other replicas are not in the ISR list, the partition remains unavailable until the leader node recovers.

  All copies in the ISR list are required, and the writing of the minimum number of copies has been verified successful. When the node where a
copy of a partition is located is faulty, the minimum number of successful copies cannot be met. As a result, service writing fails.

Configuration Impact

Evaluate reliability and performance requirements based on service scenarios and use proper parameter configuration.

NOTE

- For valuable data, you are advised to configure raid1 or raid5 for Kafka data directory disks to improve data reliability in case disk fault of a single disk.
- The **acks** parameter is named different for different Producer APIs.
  - New Producer API
    Indicates the interface defined in `org.apache.kafka.clients.producer.KafkaProducer`. The **acks** parameter name remains unchanged for this API.
  - Old Producer API
    indicates the interface defined in `kafka.producer.Producer`. The **acks** parameter is named as `request.required.acks` for this API.
- For parameters that can be modified at the topic level, the service level configurations are used by default. These parameters can be separately configured based on topic reliability requirements.
  
For example, you can configure the reliability parameters of the topic named `test`.  

```bash
kafka-topics.sh --describe --topic example-metric1 --zookeeper 192.168.1.205:2181,192.119.147.231:2181,192.119.147.228:2181 Error while executing topic command:org.apache.zookeeper.KeeperException
```

9.32 A Kafka Cluster Cannot Be Operated Using the Kafka Shell Command

Symptom

A user runs the Kafka command to view the current topic information. However, the topic information fails to be displayed.

Details are as follows:

```bash
root@Slave2bin#./kafka-topics.sh --describe --topic example-metric1 --zookeeper 192.119.147.231:2181,192.119.147.228:2181,192.119.147.227:2181 Error while executing topic command:org.apache.zookeeper.KeeperException
```

```java
```

```scala
at kafka.utils.ZkUtils.getChildren(ZkUtils.scala:798)
at kafka.admin.TopicCommand$.getTopics(TopicCommand.scala:82)
at kafka.admin.TopicCommand$.describeTopic(TopicCommand.scala:183)
at kafka.admin.TopicCommand$.main(TopicCommand.scala:66)
```
Cause Analysis

The user has checked the permission of the account for running this command. The account has the highest permission to operate the Kafka component. Therefore, this problem is not caused by permission insufficiency.

It is confirmed that the command is incorrect. The path (Znode) of the Kafka information stored on ZooKeeper should be prefixed with /kafka. The complete query command is as follows:

```bash
root@Slave2bin# ./kafka-topics.sh --describe --topic example-metric1 --zookeeper 192.119.147.231:2181,192.119.147.228:2181,192.119.147.227:2181/kafka
```

Solution

This is a command operation error. For details, see Developer Guide > Kafka Application Development > Kafka APIs > Shell Commands.

9.33 Kafka Producer Writes Oversized Records

Symptom

When a user develops a Kafka application and invokes the new interface (org.apache.kafka.clients.producer.*) as a Producer to write data to Kafka, the size of a single record is 1100055, which exceeds the value (1000012) of message.max.bytes in the Kafka configuration file server.properties. After the values of message.max.bytes and replica.fetch.max.bytes in the Kafka service configuration are changed to 5242880, the exception persists. The error information is as follows:

```
14749 [Thread-0] INFO  com.huawei.bigdata.kafka.example.NewProducer - The ExecutionException occurred : {}.
java.util.concurrent.ExecutionException: org.apache.kafka.common.errors.RecordTooLargeException: The message is 1100093 bytes when serialized which is larger than the maximum request size you have configured with the max.request.size configuration.
at org.apache.kafka.clients.producer.KafkaProducer$FutureFailure.<init>(KafkaProducer.java:739)
at org.apache.kafka.clients.producer.KafkaProducer.doSend(KafkaProducer.java:483)
at org.apache.kafka.clients.producer.KafkaProducer$FutureFailure.<init>(KafkaProducer.java:430)
at org.apache.kafka.clients.producer.KafkaProducer$FutureFailure.<init>(KafkaProducer.java:353)
at com.huawei.bigdata.kafka.example.NewProducer.run(NewProducer.java:150)
Caused by: org.apache.kafka.common.errors.RecordTooLargeException: The message is **** bytes when serialized which is larger than the maximum request size you have configured with the max.request.size configuration.
```

.......
Cause Analysis

When data is written to Kafka, the Kafka client compares the value of `max.request.size` with the size of the data to be written. If the size of the data to be written exceeds the default value of `max.request.size`, the preceding exception is reported.

Solution

Step 1  You can set the value of `max.request.size` when initializing the Kafka Producer instance.

For example, you can set this configuration item to 5252880 by referring to this instance.
```java
// Protocol type: Currently, the SASL_PLAINTEXT or PLAINTEXT protocol types can be used.
props.put(securityProtocol, kafkaProc.getValues(securityProtocol, "SASL_PLAINTEXT"));
// Service name
props.put(saslKerberosServiceName, "kafka");
props.put("max.request.size", "5252880");
```

9.34 Kafka Consumer Reads Oversized Records

Symptom

After data is written to Kafka, a user develops an application and invokes the interface `(org.apache.kafka.clients.consumer.*)` to read data from Kafka as a Consumer. However, the reading fails and the following error is reported:

```java
1687 [KafkaConsumerExample] INFO org.apache.kafka.clients.consumer.internals.AbstractCoordinator - Successfully joined group DemoConsumer with generation 1
2053 [KafkaConsumerExample] ERROR com.huawei.bigdata.kafka.example.NewConsumer - [KafkaConsumerExample], Error due to org.apache.kafka.common.errors.RecordTooLargeException: There are some messages at [Partition=Offset]: {default-0=177} whose size is larger than the fetch size 1048576 and hence cannot be ever returned. Increase the fetch size on the client (using max.partition.fetch.bytes), or decrease the maximum message size the broker will allow (using message.max.bytes).
2059 [KafkaConsumerExample] INFO com.huawei.bigdata.kafka.example.NewConsumer - [KafkaConsumerExample], Stopped
```

Cause Analysis

When reading data, the Kafka client compares the size of the data to be read with the value of `max.partition.fetch.bytes`. If the size exceeds the value of `max.partition.fetch.bytes`, the preceding exception is reported.

Solution

Step 1  When creating a Kafka Consumer instance during initialization, set `max.partition.fetch.bytes`.

For example, you can set its value to 5252880 by referring to this instance.
// Security protocol type
props.put(securityProtocol, kafkaProc.getValues(securityProtocol, "SASL_PLAINTEXT");
// Service name
props.put(saslKerberosServiceName, "kafka");
props.put("max.partition.fetch.bytes", "5252880");

----End
10.1 Invalid Hyperlink of events on the Storm UI

Issue

The hyperlink of events on the Storm UI is invalid.

Symptom

After submitting a topology, the user cannot view topology data processing logs and the events hyperlink is invalid.

Cause Analysis

The function of viewing topology data processing logs is disabled by default when an MRS cluster submits a topology.

Procedure

Step 1 Access MRS Manager.

Step 2 Choose Services > Storm. In Storm WebUI of Storm Summary, click any UI link to access the Storm WebUI.

Step 3 In the Topology Summary area, click the desired topology to view details.

Step 4 In the Topology actions area, click Kill to delete the submitted Storm topology.

Step 5 Submit the Storm topology again and enable the function of viewing topology data processing logs. Add the topology.eventlogger.executors parameter and set it to a positive integer except 0 when submitting the Storm topology. The following provides an example:

```
storm jar Path of the topology package Class name of the topology Main method Topology name -c topology.eventlogger.executors=X
```

Step 6 In the Topology Summary area on the Storm UI, click the desired topology to view details.
Step 7  In the Topology actions area, click Debug, specify the data sampling percentage, and click OK.

Step 8  Click the Spouts or Bolts task name of the topology. In Component summary, click events to view data processing logs.

![NOTE](image)

To enable the function of viewing topology data processing logs of the specified Spouts or Bolts task, click the Spouts or Bolts task name of the topology, click Debug in the Topology actions area, and enter the data sampling percentage.

--- End

10.2 Failed to Submit a Topology

Symptom

An MRS streaming cluster is installed, and ZooKeeper, Storm, and Kafka are installed in the cluster.

A topology fails to be submitted by running commands on the client.

Possible Causes

- The Storm service is abnormal.
- The client user is not authenticated or the authentication has expired.
- The storm.yaml file in the submitted topology conflicts with that on the server.

Cause Analysis

A user fails to submit the topology. The possible cause is that the client or Storm is faulty.

1. On MRS Manager, choose Services > Storm. The status of Storm is Good and the monitoring indicators show no errors.
2. Check the submission logs of the client. It is found that the error information "KeeperExceptionSessionExpireException" is printed, as shown in the following:
The preceding error occurs because security authentication is not performed before the topology is submitted or the TGT expires after authentication.

For details about the solution, see Step 1.

3. Check the client submission log. It is found that the "ExceptionInitializerError" exception information is printed, and the message "Found multiple storm.yaml resources" is displayed. The following is an example:

```
Exception is thrown "Faulty" java.lang.ExceptionInDeclarationError
at org.apache.storm.topology.TopologyBuilder.addTopology(TopologyBuilder.java:159)
Caused by: java.lang.RuntimeException: Found multiple storm.yaml resources. You're probably loading the storm jar with your topology jar.
```

This error occurs because the storm.yaml file in the service JAR package conflicts with that on the server.

For details about the solution, see Step 2.

4. If the fault is not caused by the preceding reasons, see Topology Submission Fails and the Message "Failed to check principle for keytab" Is Displayed.

Solution

Step 1 An authentication error occurs.

1. Log in to the node where the client resides and switch to the client directory.

2. Run the following command to submit the task again: (Replace the service JAR package and topology based on the site requirements.)

```bash
source bigdata_env
kinit Username
storm jar storm-starter-topologies-0.10.0.jar storm.starter.WordCountTopology test
```

Step 2 The topology package is abnormal.

Check the service JAR package, delete the storm.yaml file from the service JAR package, and submit the task again.

----End
10.3 Topology Submission Fails and the Message "Failed to check principle for keytab" Is Displayed

Symptom

An MRS streaming cluster in security mode is installed, and ZooKeeper, Storm, and Kafka are installed in the cluster.

When a topology is defined to access components such as HDFS and HBase and the topology fails to be submitted using client commands.

Possible Causes

- The submitted topology does not contain the keytab file of the user.
- The keytab file contained in the submitted topology is inconsistent with the user who submits the topology.
- The user.keytab file exists in the /tmp directory on the client, and the owner is not the running user.

Cause Analysis

1. Check the logs. Error information "Can not found user.keytab in storm.jar" is found. Details are as follows:

```java
[main] INFO  b.s.StormSubmitter - Get principle for stream@HADOOP.COM success
[main] ERROR b.s.StormSubmitter - Can not found user.keytab in storm.jar.
Exception in thread "main" java.lang.RuntimeException: Failed to check principle for keytab
at backtype.storm.StormSubmitter.submitTopologyAs(StormSubmitter.java:219)
```

Check the JAR file of the submitted topology. It is found that the keytab file is not contained.

2. Check the logs. Error information "The submit user is invalid, the principle is" is found. Details are as follows:

```java
[main] INFO  b.s.StormSubmitter - Get principle for stream@HADOOP.COM success
[main] WARN  b.s.s.a.k.ClientCallbackHandler - Could not login: the client is being asked for a password, but the client code does not currently support obtaining a password from the user. Make sure that the client is configured to use a ticket cache (using the JAAS configuration setting 'useTicketCache=true') and restart the client. If you still get this message after that, the TGT in the ticket cache has expired and must be manually refreshed. To do so, first determine if you are using a password or a keytab. If the former, run kinit in a Unix shell in the environment of the user who is running this client using the command 'kinit <princ>' (where <princ> is the name of the client's Kerberos principal). If the latter, do 'kinit -k -t <keytab> <princ>' (where <princ> is the name of the Kerberos principal, and <keytab> is the location of the keytab file). After manually refreshing your cache, restart this client. If you continue to see this message after manually refreshing your cache, ensure that your KDC host's clock is in sync with this host's clock.
```

```java
Exception in thread "main" java.lang.RuntimeException: Failed to check principle for keytab
at backtype.storm.StormSubmitter.submitTopologyAs(StormSubmitter.java:219)
```
The authenticated user used to submit the topology is stream. However, the system displays a message indicating that the submit user is invalid during topology submission, indicating that the internal verification fails.

3. Check the JAR file of the submitted topology. It is found that the keytab file is contained.

The principal parameter is set to $\text{zmk\_kafka}$ in the $\text{user.keytab}$ file.

```
[root@88-5-140-6 client]$ klist -kt user.keytab
Keytab name: FILE:user.keytab
KYN0           Timestamp                   Principal
--------------------------------------------------------------------------
  1 12/19/16 16:28:17 zmk_kafka@HADOOP.COM
  1 12/19/16 16:28:17 zmk_kafka@HADOOP.COM
```

It is found that the authenticated user does not match the principal in the $\text{user.keytab}$ file.

4. Check the logs and find the error information "Delete the tmp keytab file failed, the keytab file is:/tmp/user.keytab". The detailed information is as follows:

```
[main] WARN b.s.StormSubmitter - Delete the tmp keytab file failed, the keytab file is: /tmp/
user.keytab
[main] ERROR b.s.StormSubmitter - The submit user is invalid, the principle is: hbase1@HADOOP.COM
Exception in thread "main" java.lang.RuntimeException: Failed to check principle for keytab
at backtype.storm.StormSubmitter.submitTopologyAs(StormSubmitter.java:213)
at backtype.storm.StormSubmitter.submitTopology(StormSubmitter.java:286)
at backtype.storm.StormSubmitter.submitTopology(StormSubmitter.java:170)
at com.touchstone.storm.cmcc.CmccDataHbaseTopology.main(CmccDataHbaseTopology.java:183)
```

Check the /tmp directory. It is found that the $\text{user.keytab}$ file exists and the file owner is not the running user.

**Solution**

- Ensure that the $\text{user.keytab}$ file is carried when the topology is submitted.
- Ensure that the user for submitting the topology is the same as that of the $\text{user.keytab}$ file.
- Delete the $\text{user.keytab}$ file from the /tmp directory.

### 10.4 The Worker Log Is Empty After a Topology Is Submitted

**Symptom**

After a topology is remotely submitted in Eclipse, the detailed information about the topology cannot be viewed on the Storm web UI, and the Worker node where Bolt and Spout of each topology are located keeps changing. The Worker log is empty.

**Possible Causes**

The Worker process fails to be started, triggering Nimbus to re-allocate tasks and start the Worker process on other Supervisors. The Worker process continues to restart. As a result, the Worker node keeps changing, and the Worker log is empty. The possible causes of the Worker process startup failure are as follows:
The submitted JAR package contains the `storm.yaml` file. Storm specifies that each classpath can contain only one `storm.yaml` file. If there is more than one `storm.yaml` file, an exception occurs. Use the Storm client to submit the topology. The classpath configuration of the client is different from the classpath configuration of Eclipse. The client automatically loads the JAR package of the user to classpath. As a result, two `storm.yaml` files exist in classpath.

The initialization of the Worker process takes a long time, which exceeds the Worker startup timeout period set in the Storm cluster. As a result, the Worker process is killed and reallocated.

### Troubleshooting Process

1. Use the Storm client to submit the topology and check whether the `storm.yaml` file is duplicate.
2. Repack the JAR file and submit the topology again.
3. Modify the Worker startup timeout parameter in the Storm cluster.

### Procedure

**Step 1** If the Worker log is empty after the topology is remotely submitted using Eclipse, use the Storm client to submit the JAR package corresponding to the topology and view the prompt message.

For example, if the JAR package contains two `storm.yaml` files in different paths, the following information is displayed:

```
Exception in thread "main" java.lang.ExceptionInInitializerError
  at com.huawei.streaming.storm.example.WordCountTopology.createConf(WordCountTopology.java:132)
  at com.huawei.streaming.storm.example.WordCountTopology.remoteSubmit(WordCountTopology.java:120)
  at com.huawei.streaming.storm.example.WordCountTopology.main(WordCountTopology.java:101)
Caused by: java.lang.RuntimeException: Found multiple storm.yaml resources. You're probably bundling the Storm jars with your topology jar.
  [jar:file:/opt/huawei/fi_client/Streaming/streaming-0.9.2/bin/stormDemo.jar!/storm.yaml,
   file:/opt/huawei/fi_client/Streaming/streaming-0.9.2/conf/storm.yaml]
  at backtype.storm.utils.Utils.findAndReadConfigFile(Utils.java:151)
  at backtype.storm.utils.Utils.readStormConfig(Utils.java:206)
  at backtype.storm.utils.<init>(Utils.java:70)
```

**Step 2** Compress the JAR package again. Ensure that the package does not contain the `storm.yaml` file and JAR packages related to `log4j` and `slf4j-log4j`.

**Step 3** Use Eclipse to remotely submit the new JAR package.

**Step 4** Check whether the topology details and Worker logs can be viewed on the web UI.

**Step 5** On MRS Manager, choose **Services > Storm > Service Configuration**, modify the Worker startup timeout parameter of the Storm cluster (for details about the parameter description, see **Related Information**), save the modification, and restart the Storm service.
Step 6  Submit the JAR package to be run again.

----End

Related Information

1. The `nimbus.task.launch.secs` and `supervisor.worker.start.timeout.secs` parameters indicate the topology startup timeout tolerance of the Nimbus and supervisor, respectively. Generally, the value of `nimbus.task.launch.secs` must be greater than or equal to that of `supervisor.worker.start.timeout.secs`. It is recommended that the value of `nimbus.task.launch.secs` be slightly greater or equal to that of `supervisor.worker.start.timeout.secs`. Otherwise, the task reallocation efficiency will be affected.

   - `nimbus.task.launch.secs`: If the Nimbus does not receive the heartbeat message sent by the topology task within the period specified by this parameter, the Nimbus re-allocates the topology to another supervisor and updates the task information in ZooKeeper. The supervisor reads the task information in ZooKeeper and compares it with the topology started. If the topology does not belong to the supervisor, the supervisor deletes the metadata of the topology, that is, the `/srv/Bigdata/streaming_data/stormdir/supervisor/stormdist/{worker-id}` directory.

   - `supervisor.worker.start.timeout.secs`: After the supervisor starts a worker, if no heartbeat message is received from the worker within the period specified by this parameter, the supervisor stops the worker and waits for worker rescheduling. Generally, the value of this parameter is increased when the service startup takes a long time to ensure that the worker can be started successfully.

   If the value of `supervisor.worker.start.timeout.secs` is greater than that of `nimbus.task.launch.secs`, the worker is still started before the tolerance time of supervisor ends. However, the Nimbus considers that the service startup times out and allocates the service to another host. The background thread of the supervisor finds that the tasks are inconsistent and deletes the metadata of the topology. As a result, when the worker attempts to read `stormconf.ser` during startup, the file does not exist, and "FileNotFoundException" is thrown.
2. The `nimbus.task.timeout.secs` and `supervisor.worker.timeout.secs` parameters indicate the timeout tolerance time for the Nimbus and supervisor to report heartbeat messages during topology running. Generally, the value of `nimbus.task.timeout.secs` must be slightly greater than or equal to that of `supervisor.worker.timeout.secs`.

10.5 Worker Runs Abnormally After a Topology Is Submitted and Error "Failed to bind to:host:ip" Is Displayed

**Symptom**

After the service topology is submitted, the Worker cannot be started normally. Check the Worker log. The log records "Failed to bind to: host:ip."

**Possible Causes**

The random port range is incorrectly configured.

**Troubleshooting Process**

1. Check related information in the Worker log.
2. Check the process information about the bond port.
3. Check the random port range.

**Cause Analysis**

1. Use SSH to log in to the host where the Worker fails to be started and run the `netstat -anp | grep <port>` command to check the ID of the process that occupies the port. In the preceding command, change the port to the actual port number.
2. Run the `ps -ef | grep <pid>` command to view process details. In the command, `pid` indicates the actual process ID.
It is found that the worker process occupies the port. This process is another topology service process. According to the process details, port 29122 is allocated to the process.

3. Run the `lsof -i:<port>` command to view connection details. In the preceding command, `port` indicates the actual port number.

   ![lsof_command_output]

   It is found that port 29101 connects to port 21005 of the peer end, and port 21005 is the Kafka server port.

   It indicates that the service layer connects to Kafka to obtain messages as a client. Service ports are allocated based on the random port range of the OS.

4. Run the `cat /proc/sys/net/ipv4/ip_local_port_range` command to check the random port range.

   ![cat_command_output]

5. It is found that the random port range is too large and conflicts with the service port range of MRS.

   ![NOTE]

   The MRS service port number ranges from 20000 to 30000.

**Procedure**

**Step 1** Modify the random port range.

```
vi /proc/sys/net/ipv4/ip_local_port_range
32768 61000
```

**Step 2** Stop the service process that occupies the service port to release the port. (Stop the service topology.)

----End
10.6 "well-known file is not secure" Is Displayed When the jstack Command Is Used to Check the Process Stack

Symptom

Run the jstack command to check the process stack information. The error message "well-known file is not secure" is displayed.

```
/home/hadoop02 ~> jstack 62517
62517: well-known file is not secure
```

Cause Analysis

1. The user running the jstack command is inconsistent with the user submitting the process for viewing the pid information.
2. Storm uses the feature of differentiating users for implementing tasks. When the worker process is started, the process UID and GID are changed to the user submitting the task and ficommon. This way, logviewer can access logs of the worker process and only log file permission 640 is open. After the user is changed, the jstack and jmap commands fail to be executed for the worker process, because the default GID of the user is not ficommon. You need to run the ldap command to change the user GID to 9998 (ficommon).

Solution

You can use either of the following two methods to resolve the problem:

Method 1: View the process stack on the native Storm page.

Step 1  On the native Storm page, choose MRS Manager > Services.
Step 2 Select the topology to be viewed.

Step 3 Select the spout or bolt to be viewed.

Step 4 Select the log file of the node to be viewed, and then click **JStack** or **Heap**. **JStack** corresponds to the stack information, and **Heap** corresponds to the heap information.
Method 2: View the process stack by modifying user-defined parameters.

**Step 1** Log in to MRS Manager, choose Services > Storm > Service Configuration, and set Type to All.

**Step 2** In the navigation tree on the left, choose supervisor > Customize and add the variable supervisor.run.worker.as.user=false.

**Step 3** Click Save Configuration and select Restart the affected services or instances. Click OK to restart the services.

**Step 4** Submit the topology again.

**Step 5** Switch to the omm user on the background node and run the jps command to view the PID of the worker process.

```
omm@hadoop02:~> jps | grep worker
22455 worker
111402 worker
```

**Step 6** Run the jstack pid command to view the jstack information.
10.7 When the Storm-JDBC plug-in is used to develop Oracle write Bolts, data cannot be written into the Bolts.

**Symptom**

When the Storm-JDBC plug-in is used to develop Oracle write Bolts, the Oracle database can be connected, but data cannot be written to the Oracle database.

**Possible Causes**

- The topology definition is abnormal.
- The definition of the database table result is abnormal.

**Cause Analysis**

1. On the Storm web UI, check the DAG of the topology. The DAG is consistent with the topology definition.
2. The definition of the KeyWordFilter Bolt is consistent with the **expParser** field.

```java
@Override
public void declareOutputFields(OutputFieldsDeclarer declarer) {
    declarer.declare(new Fields("timestamp", "keyword", "hostname", "message", "kafka_topic"));
}

if (flag)
    String keyword = expParser.getKeyword();
    System.out.println("message ");
    collector.emit(new Values(timestamp, keyword, hostname, message, kafka_topic));
```

3. View the table definition in the Oracle database. The field name is in uppercase, which is inconsistent with flow definition field name.

4. When the execute method is debugged independently, it is found that the thrown field does not exist.
Procedure

The field name of the stream definition is changed to uppercase letters, which is the same as that defined in the database table.

10.8 The GC Parameter Configured for the Service Topology Does Not Take Effect

Symptom

The `topology.worker.childopts` parameter in the service topology code does not take effect. The key log is as follows:

```
Start uploading file '/opt/jar/example.jar' to '/srv/BigData/streaming/stormdir/nimbus/inbox/stormjar-8d3b778d-69ea-4f8e-ba58-01aa2036753.jar' (65574612 bytes)
[==================================================] 65574612 / 65574612
File '/opt/jar/example.jar' uploaded to '/srv/BigData/streaming/stormdir/nimbus/inbox/stormjar-8d3b778d-69ea-4f8e-ba58-01aa2036753.jar' (65574612 bytes)
[main] INFO b.s.StormSubmitter - Finished submitting topology: word-count
```

The following worker process information is displayed after the `ps -ef | grep worker` command is executed:

```
```

Cause Analysis

1. `topology.worker.gc.childopts`, `topology.worker.childopts`, and `worker.gc.childopts` (server parameters) have priorities:

2. If the client parameter `topology.worker.childopts` is set, this parameter and the server parameter `worker.gc.childopts` are configured together. However, for two same parameters, one of them will be overwritten by the other parameter after it. Take parameter `-Xmx`, as shown in the red box of the preceding figure, as an example, parameter `-Xmx1G` overwrites `-Xmx4096m`.

3. If parameter `topology.worker.gc.childopts` is configured on the client, the parameter `worker.gc.childopts` on the server will be replaced.
Solution

Step 1 If you want to modify the JVM parameter of the topology, you can directly modify the `topology.worker.gc.childopts` parameter in the command or modify the parameter on the server. When `topology.worker.gc.childopts` is set to `-Xms4096m -Xmx4096m -XX:+UseG1GC -XX:+PrintGCDetails -XX:+PrintGCDateStamps -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=1M:

```
[main-SendThread(187.7.61.88:24002)] INFO o.a.s.s.o.a.z.ClientCnxn - Socket connection established, initiating session, client: /187.7.61.88:44694, server: 187.7.61.88/187.7.61.88:24002
[main-SendThread(187.7.61.88:24002)] INFO o.a.s.s.o.a.z.ClientCnxn - Session establishment complete on server 187.7.61.88/187.7.61.88:24002, sessionid = 0x16037a6e5f092575, negotiated timeout = 40000
[main-EventThread] INFO o.a.s.s.o.a.c.f.s.ConnectionStateManager - State change: CONNECTED
[main] INFO b.s.u.StormBoundedExponentialBackoffRetry - The baseSleepTimeMs [1000] the maxSleepTimeMs [10000] the maxRetries [1]
[main] INFO o.a.s.s.o.a.z.Login - successfully logged in.
[main-EventThread] INFO o.a.s.s.o.a.z.ClientCnxn - EventThread shut down for session: 0x16037a6e5f092575
[main] INFO o.a.s.s.o.a.z.ZooKeeper - Session: 0x16037a6e5f092575 closed
[main] INFO b.s.StormSubmitter - Uploading topology jar /opt/jar/example.jar to assigned location: /srv/BigData/streaming/stormdir/nimbus/inbox/stormjar-86855b6b-133e-478d-b415-fa96e63e553f.jar
Start uploading file '/opt/jar/example.jar' to '/srv/BigData/streaming/stormdir/nimbus/inbox/stormjar-86855b6b-133e-478d-b415-fa96e63e553f.jar' (74143745 bytes)
[main] INFO b.s.StormSubmitter - Successfully uploaded topology jar to assigned location: /srv/BigData/streaming/stormdir/nimbus/inbox/stormjar-86855b6b-133e-478d-b415-fa96e63e553f.jar (74143745 bytes)
file '/opt/jar/example.jar' uploaded to '/srv/BigData/streaming/stormdir/nimbus/inbox/stormjar-86855b6b-133e-478d-b415-fa96e63e553f.jar' (74143745 bytes)
[main] INFO b.s.StormSubmitter - Successfully uploaded topology jar to assigned location: /srv/BigData/streaming/stormdir/nimbus/inbox/stormjar-86855b6b-133e-478d-b415-fa96e63e553f.jar (74143745 bytes)
```

Step 2 Run the `ps -ef | grep worker` command to view the worker process information:

```
```

10.9 Internal Server Error Is Displayed When the User Queries Information on the UI

Symptom

An MRS cluster is installed, and ZooKeeper and Storm are installed in the cluster. "Internal Server Error" is displayed when a user accesses information from the Storm Status page of MRS Manager.

The detailed information is as follows:

```
org.apache.thrift7.transport.TTransportException: Frame size (306030) larger than max length (1048576)
```

----End
Possible Causes

- Nimbus of Storm is abnormal.
- Storm cluster information exceeds the default Thrift transmission size.

Cause Analysis

1. On the MRS Manager page, choose Services > Storm. The status of Storm is Good and the monitoring indicators show no errors.

   **Figure 10-2 Storm service status**

   ![Storm service status](image)

2. Click the Instance tab and check the status of the Nimbus instance. The status is normal.

   **Figure 10-3 Nimbus instance status**

   ![Nimbus instance status](image)

3. Click the Service Configuration tab and check the Thrift configuration of the Storm cluster. It is found that nimbus.thrift.max_buffer_size is set to 1048576 (1 MB).

   **Figure 10-4 Configuration of nimbus.thrift.max_buffer_size**

   ![Configuration of nimbus.thrift.max_buffer_size](image)
4. The preceding configuration is the same as that in the exception information, indicating that the buffer size of Thrift is less than that required by the cluster information.

Procedure

Adjust the Thrift buffer size of the Storm cluster.

**Step 1** On MRS Manager, choose Services > Storm > Service Configuration and set Type to All.

**Step 2** Change the value of `nimbus.thrift.max_buffer_size` to 10485760 (10 MB).

**Step 3** Click Save Configuration and select Restart the affected services or instances. Click OK to restart the services.

----End
11.1 Class Cannot Be Found After Flume Submits Jobs to Spark Streaming

Issue
After Flume submits jobs to Spark Streaming, the class cannot be found.

Symptom
After the Spark Streaming code is packed into a JAR file and submitted to the cluster, an error message is displayed indicating that the class cannot be found. The following two methods are not useful:

1. When submitting a Spark job, run the --jars command to reference the JAR file of the class.
2. Import the JAR file where the class resides to the JAR file of Spark Streaming.

Cause Analysis
Some JAR files cannot be loaded during Spark job execution, resulting that the class cannot be found.

Procedure

Step 1 Run the --jars command to load the flume-ng-sdk-{version}.jar dependency package.

Step 2 Modify the two configuration items in the spark-default.conf file:

spark.driver.extraClassPath=$PWD/*: {Add the original value}
spark.executor.extraClassPath =$PWD/*

Step 3 Run the job successfully. If an error is reported, check which JAR is not loaded and perform step 1 and step 2 again.

----End
11.2 Failed to Install a Flume Client

Symptom
A Flume client fails to be installed, and "JAVA_HOME is null" or "flume has been installed" is displayed.

CST 2016-08-31 17:02:51 [flume-client install]: JAVA_HOME is null in current user, please install the JDK and set the JAVA_HOME
CST 2016-08-31 17:02:51 [flume-client install]: check environment failed.
CST 2016-08-31 17:02:51 [flume-client install]: check param failed.
CST 2016-08-31 17:02:51 [flume-client install]: install flume client failed.

Cause Analysis
- Environment variables are checked during Flume client installation. If no Java is available, an error message is displayed stating "JAVA_HOME is null" and the installation quits.
- If Flume has been installed in the specified directory, an error message is displayed stating "flume has been installed" during client installation and the installation quits.

Solution

Step 1 Run the following command if an error message is displayed stating "JAVA_HOME is null":

```bash
export JAVA_HOME=Java path
```
Set JAVA_HOME and execute the installation script again.

Step 2 If a Flume client has been installed under the specified directory, uninstall the client and use another directory.

----End

11.3 A Flume Client Cannot Connect to the Server

Symptom
A user installs a Flume client and sets an Avro sink to communicate with the server. However, the Flume server cannot be connected.

Cause Analysis
1. The server is incorrectly configured and the monitoring port fails to be started up. For example, an incorrect IP address or an occupied port is configured for the Avro source of the server. View Flume run logs.
2. If encrypted transmission is used, the certificate or password is incorrect.

3. The network connection between the client and the server is abnormal.

Solution

Step 1 Set a correct IP address (an IP address of the local host). If the port has been occupied, configure another free port.

Step 2 Configure a correct certificate path.

Step 3 Contact the network administrator to restore the network.

11.4 Flume Data Fails to Be Written to the Component

Symptom

After the Flume process is started, Flume data cannot be written to the corresponding component. (The following uses writing data from the server to HDFS as an example.)

Cause Analysis

1. HDFS is not started or is faulty. View Flume run logs.

2. The HDFS sink is not started. Check the Flume run log. It is found that the Flume current metrics file does not contain sink information.
11.5 Flume Server Process Fault

Symptom

After Flume runs for a period of time, the Flume instance is in the faulty state on MRS Manager.

Cause Analysis

If the Flume file or folder permission is abnormal, the following information is displayed on MRS Manager after the restart:

```
[2019-02-26 13:38:02]RoleInstance prepare to start failure [(ScriptExecutionResult=ScriptExecutionResult
[exitCode=126, output=, enMsg=sh: line 1: /opt/Bigdata/MRS_XXX/install/FusionInsight-Flume-1.6.0/
flume/bin/flume-manage.sh: Permission denied
```

Solution

Compare the file and folder permissions with those for the Flume node that is running properly and correct the file or folder permissions.

11.6 Flume Data Collection Is Slow

Symptom

After Flume is started, it takes a long time for Flume to collect data.

Cause Analysis

1. The heap memory of Flume is not properly set. As a result, the Flume process keeps in the GC state. View Flume run logs.

```
2.5817610 secs] 4153654K->3843458K(4160256K), [CMS Perm : 27335K->27335K(45592K),2.5820080
SECS] [Times: user=2.63, sys0.00, real=2.59 secs]
```
2. The deletePolicy policy configured for the Spooldir source is **immediate**.

**Solution**

**Step 1** Increase the size of the heap memory (xmx).

**Step 2** Change the deletePolicy policy of the Spooldir source to **never**.

----End

### 11.7 Flume Startup Failure

#### Symptom

The Flume service fails to be installed or restarted.

#### Cause Analysis

1. The heap memory of Flume is greater than the remaining memory of the server. Check the Flume startup log.

   ```
   [CST 2019-02-26 13:31:43][INFO] [[checkMemoryValidity:124]] [GC_OPTS is invalid: Xmx(40960000MB) is bigger than the free memory(56118MB) in system.] [9928]
   ```

2. The permission on the Flume file or folder is abnormal. The following information is displayed on the GUI or in the background:

   ```
   [2019-02-26 13:38:02]RoleInstance prepare to start failure
   [ScriptExecutionResult=ScriptExecutionResult [exitCode=126, output=, errMsg=sh: line 1: /opt/Bigdata/MRS_XXX/install/FusionInsight-Flume-1.6.0/bin/flume-manage.sh: Permission denied]
   ```

3. If JAVA_HOME is incorrectly configured, check the Flume agent startup log.

   ```
   Info: Sourcing environment configuration script /opt/FlumeClient/fusioninsight-flume-1.6.0/conf/flume-env.sh
   *
   *'[-n '' ']
   + exec /tmp/MRS-Client/MRS_Flume_ClientConfig/JDK/jdk-8u18/bin/java -XX:OnOutOfMemoryError=bash /opt/FlumeClient/fusioninsight-flume-1.6.0/bin/out_memory_error.sh /opt/FlumeClient/fusioninsight-flume-1.6.0/conf %9p -Xms2G -Xmx4G -XX:CMSFullGCsBeforeCompaction=1 -XX:+UseConcMarkSweepGC -XX:+CMSParallelRemarkEnabled -XX:+UseCMSCompactAtFullCollection -Xerberos.domain.name=hadoop.hadoop.com -verbose:gc -XX:+UseGCLargeAllocation -XX:NumberOfGCLargeAllocations=10 -XX:GCLargeAllocationSize=1M -XX:+PrintGCDetails -XX:+PrintGCDateStamps -Xloggc:/var/log/Bigdata/flume-client-1/flume/log/flume-root-20190226134231%9p-gc.log -Dproc.org.apache.flume.node.Application -Dproc.name=client -Dproc.conf_file=/opt/FlumeClient/fusioninsight-flume-1.6.0/conf/properties.properties -Djava.security.krb5.conf=/opt/FlumeClient/fusioninsight-flume-1.6.0/conf/kerberos/karber5.conf -Dflume.instance.id=884174180 -Dflume.agent.name=clientName1 -Dflume.role=client -Dlog4j.configuration=log4j.properties -Djava.security.auth.login.config=/opt/FlumeClient/fusioninsight-flume-1.6.0/conf/kerberos/jaas.conf -Dflume.called.from.service -Dflume.conf.dir=/opt/FlumeClient/fusioninsight-flume-1.6.0/conf -Dflume.metric.conf.dir=/opt/FlumeClient/fusioninsight-flume-1.6.0/conf -Dflume.service.id=flume-client-1 -Dflume.application.home.path=/opt/FlumeClient/fusioninsight-flume-1.6.0/conf -Dflume.client.name=clientName1
   ```

**Solution**

**Step 1** Increase the value of heap memory (xmx).
Step 2 Compare the file and folder permissions with those for the Flume node that is started properly and modify the incorrect file or folder permissions.

Step 3 Reconfigure JAVA_HOME. On the client, replace the value of JAVA_HOME in the $ {install_home}/fusioninsight-flume-1.6.0/conf/ENV_VARS file. On the server, replace the value of JAVA_HOME in the ENV_VARS file in the etc directory of the Flume.

**NOTE**

$\text{install\_home}$ is the installation path of the Flume client.

----End
12.1 A Job Is Running on Hue

Issue
The customer finds that a job is running on Hue.

Symptom
After the customer's MRS is installed, the job is running on Hue but the running job is not operated by the customer.

Cause Analysis
This job is a permanent job generated when the system connects to JDBC after Spark is started.

Procedure
This is not a problem. No handling is required.
12.2 HQL Fails to Be Executed on Hue Using Internet Explorer

Symptom
Using Internet Explorer to access Hive Editor and execute all HQL statements on Hue fails and the system prompts "There was an error with your query".

Cause Analysis
Internet Explorer has functional problems and cannot process AJAX POST requests containing form data in 307 redirection. Use a compatible browser.

Solution
Use Google Chrome 21 or later.

12.3 Hue (Active) Cannot Open Web Pages

Symptom
The following information is displayed on the web UI of Hue (active):

Service Unavailable
The server is temporarily unable to service your request due to maintenance downtime or capacity problems. Please try again later.

Cause Analysis
- The Hue configuration has expired.
- The configuration of the Hue service needs to be modified manually in a single-Master cluster of MRS 2.0.1 or later.

Solution
- If the Hue configuration has expired, restart the Hue service.
- Manually modify the Hue service configuration for a single-Master cluster of MRS 2.0.1 or later.
  a. Log in to the Master node.
  b. Run the hostname -i command to obtain the IP address of the local host.
  c. Run the following command to obtain the value of HUE_FLOAT_IP:
     grep "HUE_FLOAT_IP" /opt/Bigdata/MRS_Current/1_*/etc*/ENV_VARS
  d. Check whether the local IP address is the same as the value of HUE_FLOAT_IP. If they are different, change the value of HUE_FLOAT_IP to the local IP address.
  e. Restart the Hue service.
12.4 Failed to Access the Hue Web UI

Issue
An error page is displayed when the Hue web UI is accessed.

Symptom
The following error information is displayed on the Hue web UI:

503 Service Unavailable
The server is temporarily unable to service your request due to maintenance downtime or capacity problems. Please try again later.

Cause Analysis
- The Hue configuration has expired.
- The configuration of the Hue service needs to be modified manually in a single-Master cluster of MRS 2.0.1 or later.

Procedure

Step 1 Log in to the Master node.
Step 2 Run the `hostname -i` command to obtain the IP address of the local host.
Step 3 Run the following command to obtain the value of `HUE_FLOAT_IP`:

```
grep "HUE_FLOAT_IP" /opt/Bigdata/MRS_Current/1_*/etc*/ENV_VARS
```
Step 4 Check whether the local IP address is the same as the value of `HUE_FLOAT_IP`. If they are different, change the value of `HUE_FLOAT_IP` to the local IP address.
Step 5 Restart the Hue service.

-----End

12.5 HBase Tables Cannot Be Loaded on the Hue Web UI

Issue
After Hive data is imported to HBase on the Hue page, an error message is displayed, indicating that the HBase table cannot be detected.

Symptom
In the Kerberos cluster, the IAM sub-account does not have sufficient permissions. As a result, the HBase table cannot be loaded.
Cause Analysis

The IAM subaccount does not have sufficient permissions.

Procedure

Step 1  Log in to MRS Manager.
Step 2  Choose System > Manage User.
Step 3  Locate the row that contains the target user, and click Modify.
Step 4  Add the user to the supergroup group.
Step 5  Click OK. The modification is complete.

----End

Summary and Suggestions

If Kerberos authentication is enabled for a cluster, "No data available" is displayed on the page. In this case, check the permission first.
13.1 All NameNodes Become the Standby State After the NameNode RPC Port of HDFS Is Changed

Issue

After the NameNode RPC port is changed on the page and HDFS is restarted, all NameNodes are in the standby state, causing a cluster exception.

Symptom

All NameNodes are in the standby state, causing a cluster exception.

Possible Cause

After the cluster is installed and started, if the NameNode RPC port is changed, the Zkfc service must be formatted to update node information on ZooKeeper.

Procedure

Step 1 Log in to the MRS management console. On the Active Clusters page, click the cluster name. The cluster details page is displayed.

Step 2 Click View next to Cluster Manager to access the MRS Manager page.

Step 3 On the MRS Manager page, choose Services > HDFS and click Stop.

NOTE

Do not stop related services when stopping HDFS.

Step 4 After the services are stopped, log in to the Master node whose RPC port is changed.

NOTE

If the RPC port is changed on both Master nodes, you can log in to either of the Master nodes.
Step 5 Run the `su - omm` command to switch to user `omm`.

**NOTE**

For a security cluster, run the `kinit hdfs` command for authentication.

Step 6 Run the following command to load the environment variable script to the environment:

```bash
cd ${BIGDATA_HOME}/MRS_X.X.X/1_8_Zkfc/etc
source ${BIGDATA_HOME}/MRS_X.X.X/install/FusionInsight-Hadoop-3.1.1/hadoop/sbin/exportENV_VARS.sh
```

**NOTE**

In the preceding command, `MRS_X.X.X` and `1_8` vary depending on the actual version.

Step 7 After the loading is complete, run the following command to format the Zkfc:

```bash
cd ${HADOOP_HOME}/bin
./hdfs zkfc -formatZK
```

Step 8 After the formatting is successful, restart HDFS on MRS Manager.

**NOTE**

If the RPC port of the NameNode is changed, the configuration file must be updated for all clients that have been installed.

----End

### 13.2 An Error Is Reported When the HDFS Client Is Used After the Host Is Connected Using a Public Network IP Address

**Issue**

When the host is connected using a public network IP address, the HDFS client cannot be used and the message "`-bash: hdfs: command not found`" is displayed when the HDFS is running.

**Symptom**

When the host is connected using a public network IP address, the HDFS client cannot be used and the message "`-bash: hdfs: command not found`" is displayed when the HDFS is running.

**Possible Cause**

The environment variables are not set before the user logs in to the Master node and runs the command.
Procedure

Step 1  Log in to any Master node as user root.

Step 2  Run the source /opt/client/bigdata_env command to configure environment variables.

Step 3  Run the hdfs command to use the HDFS client.

13.3 Failed to Use Python to Remotely Connect to the Port of HDFS

Issue

Failed to use Python to remotely connect to the port of HDFS.

Symptom

Failed to use Python to remotely connect to port 50070 of HDFS.

Cause Analysis

The default port of open source HDFS is 50070 for versions earlier than 3.0.0 and is 9870 for version 3.0.0 or later. The port used by the user does not match the HDFS version.

Step 1  Log in to the active Master node in the cluster.

Step 2  Run the su - omm command to switch to user omm.

Step 3  Run the /opt/Bigdata/om-0.0.1/sbin/queryVersion.sh command to check the HDFS version in the cluster.

Determine the port number of the open source component based on the version number. Obtain the port number corresponding to the HDFS version. For details about how to query the port number of an open source component, see List of Open Source Component Ports.

Step 4  Run the netstat -an|grep $port command to check whether the default port number of the component exists.

If it does not exist, the default port number is changed. Change the port to the default port and reconnect to HDFS.

If it exists, contact technical support.

NOTE

- $port: indicates the default port number corresponding to the component version.
- If you have changed the default port number, use the new port number to connect to HDFS. You are advised not to change the default port number.

----End
13.4 HDFS Capacity Usage Reaches 100%, Causing Unavailable Upper-layer Services Such as HBase and Spark

**Issue**

The HDFS capacity usage of the cluster reaches 100%, and the HDFS service status is read-only. As a result, upper-layer services such as HBase and Spark are unavailable.

**Symptom**

The HDFS capacity usage is 100%, the disk capacity usage is only about 85%, and the HDFS service status is read-only. As a result, upper-layer services such as HBase and Spark are unavailable.

**Possible Cause**

Currently, NodeManager and DataNode share data disks. By default, MRS reserves 15% of data disk space for non-HDFS. You can change the percentage of data disk space by setting the HDFS parameter `dfs.datanode.du.reserved.percentage`.

If the HDFS disk usage is 100%, you can set `dfs.datanode.du.reserved.percentage` to a smaller value to restore services and then expand disk capacity.

**Procedure**

**Step 1** Log in to any Master node in the cluster.

**Step 2** Run the `source /opt/client/bigdata_env` command to initialize environment variables.

**NOTE**

If it is a security cluster, run the `kinit -kt <keytab file> <principal name>` command for authentication.

**Step 3** Run the `hdfs dfs -put ./startDetail.log /tmp` command to check whether HDFS fails to write files.

```
19/05/12 10:07:32 WARN hdfs.DataStreamer: DataStreamer Exception org.apache.hadoop.ipc.RemoteException(java.io.IOException): File /tmp/startDetail.log._COPYING_ could only be replicated to 0 nodes instead of minReplication (=1). There are 3 datanode(s) running and no node(s) are excluded in this operation.
```

**Step 4** Run the `hdfs dfsadmin -report` command to check the used HDFS capacity. The command output shows that the HDFS capacity usage has reached 100%.

```
Configured Capacity: 5389790579100 (4.90 TB)
Present Capacity: 5067618628404 (4.61 TB)
DFS Remaining: 133350196 (127.17 MB)
DFS Used: 5067485278208 (4.61 TB)
DFS Used%: 100.00%
Under replicated blocks: 10
Blocks with corrupt replicas: 0
```
Step 5  When the HDFS capacity usage reaches 100%, change the percentage of data disk space by setting the HDFS parameter `dfs.datanode.du.reserved.percentage`.

1. Log in to MRS Manager and choose Services > HDFS > Service Configuration.
2. Set Type to All, and search for `dfs.datanode.du.reserved.percentage` in the search box.
3. Change the value of this parameter to 10.

Step 6  After the modification, increase the number of disks of the Core node. For details, see Adding a New Disk to an MRS Cluster.

13.5 An Error Is Reported During HDFS and Yarn Startup

Issue

An error is reported during HDFS and Yarn startup.

Symptom

HDFS and Yarn fail to be started. The following error information is displayed: `/dev/null Permission denied`.

Possible Cause

The customer changed the permission value of `/dev/null` of the VM system to 775.
Procedure

Step 1 Log in to any Master node in a cluster as user root.

Step 2 After successful login, run the chmod 666 /dev/null command to modify the permission value of /dev/null to 666.

Step 3 Run the ls -al /dev/null command to check whether the new permission value of /dev/null is 666. If it is not, change the value to 666.

Step 4 After the modification is successful, restart HDFS and Yarn.

13.6 HDFS Permission Setting Error

Issue

When using MRS, a user can delete or create files in another user's HDFS directory.

Symptom

When using MRS, a user can delete or create files in another user's HDFS directory.

Cause Analysis

The user configured by the customer has the permission of the ficommon group, that is, the superuser group. Therefore, the user can perform any operation on HDFS. You need to remove the user's ficommon group permission.

Procedure

Step 1 Log in to the Master node in the cluster as user root.

Step 2 Run the id ${Username} command to check whether the user has the ficommon group permission.

If the user has the ficommon group permission, go to Step 3. If the user does not have the ficommon group permission, contact technical support.

NOTE

${Username}: indicates the name of the user whose HDFS permission is incorrectly set.

Step 3 Run the gpasswd -d ${Username} ficommon command to delete the user's ficommon group permission.

NOTE

${Username}: indicates the name of the user whose HDFS permission is incorrectly set.

Step 4 After the command has been executed successfully, log in to MRS Manager and choose Services > HDFS > Service Configuration.
Step 5  Set **Type** to **All**, enter `dfs.permissions.enabled` in the search box, and change the parameter value to **true**.

Step 6  After the modification, click **Save Configuration** and restart HDFS.

----End

13.7 A DataNode of HDFS Is Always in the Decommissioning State

**Issue**

A DataNode of HDFS is always in the Decommissioning state.

**Symptom**

A DataNode of HDFS fails to be decommissioned (or the Core node fails to be scaled in), but the DataNode remains in the Decommissioning state.

**Cause Analysis**

During the decommissioning of a DataNode (or scale-in of the Core node) in HDFS, the decommissioning or scale-in task fails and the blacklist is not cleared because the Master node is restarted or the NodeAgent process exits unexpectedly. In this case, the DataNode is always in the decommissioning state. The blacklist needs to be cleared manually.

**Procedure**

**Step 1**  Log in to MRS Manager and choose **Services > HDFS > Instance**.

**Step 2**  Check the HDFS service instance status, find the DataNode that is always in the decommissioning state, and copy the IP address of the DataNode.

**Step 3**  Log in to the Master1 node and run the `cd /opt/Bigdata/MRS_*/*_NameNode/etc/` command to go to the blacklist directory.

**Step 4**  Run the `sed -i "/^IPS/d" excludeHosts` command to clear the faulty DataNode information from the blacklist. Replace the IP address in the command with the IP address of the faulty DataNode queried in **Step 2**. The IP address cannot contain spaces.

**Step 5**  If there are two Master nodes, perform **Step 3** and **Step 4** on Master2.

**Step 6**  Run the following commands on the Master1 node to initialize the environment variables:

```
source /opt/client/bigdata_env
```

**Step 7**  If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If Kerberos authentication is disabled for the current cluster, skip this step:

```
kinit MRS cluster user
```
Example: `kinit admin`

**Step 8** Run the following command on the Master1 node to update the HDFS blacklist:

```
hdfs dfsadmin -refreshNodes
```

**Step 9** Run the `hdfs dfsadmin -report` command to check the status of each DataNode. Ensure that the DataNode corresponding to the IP address obtained in has been restored to the Normal state.

**Figure 13-1 DataNode status**

![DataNode status](image)

**Step 10** Log in to MRS Manager and choose Services > HDFS > Instance.

**Step 11** Select the DataNode instance that is always in the decommissioning state and choose More > Restart Instance.

**Step 12** Wait until the restart is complete and check whether the DataNode is restored.

----End

**Summary and Suggestions**

Do not perform high-risk operations, such as restarting nodes, during decommissioning (or scale-in).

**Related Information**

None

**13.8 HDFS Startup Failure Due to Insufficient Memory**

**Symptom**

After the HDFS service is restarted, HDFS is in the Bad state, the NameNode instance status is abnormal, and the system cannot exit the security mode for a long time.
**Cause Analysis**

1. In the NameNode run log (/var/log/Bigdata/hdfs/nn/hadoop-omm-namendoe-XXX.log), search for **WARN**. It is found that GC takes 63 seconds.

```
2017-01-22 14:52:32,641 | WARN | org.apache.hadoop.util.JvmPauseMonitor$Monitor@1b39fd82 | Detected pause in JVM or host machine (eg GC): pause of approximately 63750ms
GC pool 'ParNew' had collection(s): count=1 time=0ms
GC pool 'ConcurrentMarkSweep' had collection(s): count=1 time=63924ms | JvmPauseMonitor.java:189
```

2. Analyze the NameNode log /var/log/Bigdata/hdfs/nn/hadoop-omm-namendoe-XXX.log. It is found that the NameNode is waiting for block reporting and the total number of blocks is too large. In the following example, the total number of blocks is 36.29 million.

```
2017-01-22 14:52:32,641 | INFO | IPC Server handler 8 on 25000 | STATE* Safe mode ON.
The reported blocks 29715437 needs additional 6542184 blocks to reach the threshold 0.9990 of total blocks 36293915.
```

3. Check the **GC_OPTS** parameter of the NameNode:

   ![Figure 13-2 Checking the GC_OPTS parameter of the NameNode](image)

4. For details about the mapping between the NameNode memory configuration and data volume, see Table 13-1.

<table>
<thead>
<tr>
<th>Number of File Objects</th>
<th>Reference Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,000,000</td>
<td>-Xms6G -Xmx6G -XX:NewSize=512M -XX:MaxNewSize=512M</td>
</tr>
<tr>
<td>20,000,000</td>
<td>-Xms12G -Xmx12G -XX:NewSize=1G -XX:MaxNewSize=1G</td>
</tr>
<tr>
<td>50,000,000</td>
<td>-Xms32G -Xmx32G -XX:NewSize=3G -XX:MaxNewSize=3G</td>
</tr>
<tr>
<td>100,000,000</td>
<td>-Xms64G -Xmx64G -XX:NewSize=4G -XX:MaxNewSize=6G</td>
</tr>
<tr>
<td>200,000,000</td>
<td>-Xms96G -Xmx96G -XX:NewSize=8G -XX:MaxNewSize=9G</td>
</tr>
<tr>
<td>300,000,000</td>
<td>-Xms164G -Xmx164G -XX:NewSize=12G -XX:MaxNewSize=12G</td>
</tr>
</tbody>
</table>
Solution

Step 1 Modify the NameNode memory parameter based on the specifications. If the number of blocks is 36 million, change the parameter value to -Xms32G -Xmx32G -XX:NewSize=2G -XX:MaxNewSize=3G.

Step 2 Restart a NameNode and check that the NameNode can be started normally.

Step 3 Restart the other NameNode and check that the page status is restored.

----End

13.9 A Large Number of Blocks Are Lost in HDFS due to the Time Change Using ntpdate

Symptom

1. A user uses ntpdate to change the time for a cluster that is not stopped. After the time is changed, HDFS enters the safe mode and cannot be started.

2. After the system exits the safe mode and starts, about 1 TB data is lost during the hfck check.

Cause Analysis

1. A large number of blocks are lost on the native NameNode page.

Figure 13-3 Block loss
2. DataNode information on the native page shows that the number of displayed DataNode nodes is 10 less than that of actual DataNode nodes.

![Figure 13-4 Checking the number of DataNodes](image)

### Summary

Security is on. 
Satnode is off. 
14442 files and directories, 13907 blocks = 28340 total filesystem object(s). 
Heap Memory used 495.63 MB of 1.99 GB Heap Memory. Max Heap Memory is 3.98 GB. 
Non Heap Memory used 104.5 MB of 107.94 MB Committed. Non Heap Memory. Max Non Heap Memory is 1.36 GB.

<table>
<thead>
<tr>
<th>Configured Capacity</th>
<th>112.99 GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>DFS Used:</td>
<td>15.33 GB (13.68%)</td>
</tr>
<tr>
<td>Non DFS Used:</td>
<td>18.56 GB</td>
</tr>
<tr>
<td>DFS Remaining:</td>
<td>78.2 GB (69.77%)</td>
</tr>
<tr>
<td>Block Pool Used:</td>
<td>15.33 GB (13.68%)</td>
</tr>
<tr>
<td>DataNodes usage% Max/Med/Min/Max/stdDev:</td>
<td>13.56% / 13.73% / 13.73% / 0.08%</td>
</tr>
</tbody>
</table>

| Live Nodes         | 3 (Decommissioned: 0) |
| Dead Nodes         | 0 (Decommissioned: 0) |
| Decommissioning Nodes | 0 |

3. Check the DataNode run log file `/var/log/Bigdata/hdfs/dn/hadoop-omm-datanode-host name.log`. The following error information is displayed:

Major error information: Clock skew too great

![Figure 13-5 DataNode run log error](image)
Solution

**Step 1** Change the time of the 10 DataNodes that cannot be viewed on the native page.

**Step 2** On MRS Manager, restart the DataNode instances.

---End

### 13.10 CPU Usage of a DataNode Reaches 100% Occasionally, Causing Node Loss (SSH Connection Is Slow or Fails)

**Symptom**

The CPU usage of DataNodes is close to 100% occasionally, causing node loss.

#### Cause Analysis

1. A lot of write failure logs exist on DataNodes.

![DataNode write failure log](image)

2. A large number of files are written in a short time, causing insufficient DataNode memory.

![Insufficient DataNode memory](image)
**Solution**

**Step 1** Check DataNode memory configuration and whether the remaining server memory is sufficient.

**Step 2** Increase DataNode memory and restart the DataNode.

--- End

### 13.11 Manually Performing Checkpoints When a NameNode Is Faulty for a Long Time

**Symptom**

If the standby NameNode is faulty for a long time, a large amount of editlog will be accumulated. In this case, if the HDFS or active NameNode is restarted, the active NameNode reads a large amount of unmerged editlog. As a result, the HDFS or active NameNode takes a long time to restart and even fails to restart.

**Cause Analysis**

The standby NameNode periodically combines editlog files and generates the fsimage file. This process is called checkpoint. After the fsimage file is generated, the standby NameNode transfers it to the active NameNode.

**NOTE**

As the standby NameNode periodically combines editlog files, it cannot combine them when it becomes abnormal. As a result, the active NameNode needs to load many editlog files during its next startup, which occupies much memory and takes a long time.

The period of metadata combination is determined by the following parameters. If the NameNode runs for 30 minutes or one million counts of operations are performed, the checkpoint is implemented.

- `dfs.namenode.checkpoint.period`: specifies the checkpoint period. The default value is `1800s`.
- `dfs.namenode.checkpoint.txns`: specifies the times of operations for triggering the checkpoint execution. The default value is `1000000`.

**Solution**

Before restarting the HDFS or active NameNode, perform checkpoint manually to merge metadata of the active NameNode.

**Step 1** Stop workloads.

**Step 2** Obtain the host name of the active NameNode.

**Step 3** Run the following commands on the client:

```
source /opt/client/bigdata_env
kinit admin
```
Step 4 Run the following command to enable the safe mode for the active NameNode (replace linux22 with the host name of the active NameNode):

```
$ hdfs dfsadmin -fs linux22:25000 -safemode enter
```

Step 5 Run the following command to merge editlog on the active NameNode:

```
$ hdfs dfsadmin -fs linux22:25000 -saveNamespace
```

Step 6 Run the following command to make the active NameNode exit the safe mode:

```
$ hdfs dfsadmin -fs linux22:25000 -safemode leave
```

Step 7 Check whether the combination is complete.

```
$ cd /srv/BigData/namenode/current
```

Check whether the time of the first generated fsimage is the current time. If yes, the combination is complete.

13.12 Common File Read/Write Faults

Symptom

When a user performs a write operation on HDFS, the message "Failed to obtain enough extents:expected..." is displayed.

Cause Analysis

- The data receiver of the DataNode is unavailable.

The DataNode log is as follows:

```
2016-03-17 18:51:44,721 | WARN | org.apache.hadoop.hdfs.server.datanode.DataXceiverServer| DataXceiverServer@5386659f | hadoopc1h2:25009:DataXceiverServer: | DataXceiverServer.java:158
java.io.IOException: Xceiver count 4097 exceeds the limit of concurrent xcievers: 4096
```

MapReduce Service
Troubleshooting Guide

13 Using HDFS

Issue 01 (2020-07-15) Copyright © Huawei Technologies Co., Ltd. 217
- The disk space configured for the DataNode is insufficient.
- DataNode heartbeats are delayed.

**Solution**

- If the DataNode data receiver is unavailable, add the value of the HDFS parameter `dfs.datanode.max.transfer.threads` on MRS Manager.
- If disk space or CPU resources are insufficient, add DataNodes or ensure that disk space and CPU resources are available.
- If the network is faulty, ensure that the network is available.

### 13.13 Maximum Number of File Handles Is Set to a Too Small Value, Causing File Reading and Writing Exceptions

**Symptom**

The maximum number of file handles is set to a too small value, causing insufficient file handles. Writing files to HDFS is slow or file writing fails.

**Cause Analysis**

1. The DataNode log `/var/log/Bigdata/hdfs/dn/hadoop-omm-datanode-XXX.log` contains exception information "java.io.IOException: Too many open files."

```
java.io.IOException: Too many open files
at sun.nio.ch.ServerSocketChannelImpl.accept0(Native Method)
at sun.nio.ch.ServerSocketChannelImpl.accept(ServerSocketChannelImpl.java:241)
at sun.nio.ch.ServerSocketAdaptor.accept(ServerSocketAdaptor.java:100)
at org.apache.hadoop.hdfs.net.TcpPeerServer.accept(TcpPeerServer.java:134)
at org.apache.hadoop.hdfs.server.datanode.DataXceiverServer.run(DataXceiverServer.java:137)
at java.lang.Thread.run(Thread.java:745)
```

2. The error indicates insufficient file handles. File handles cannot be opened and data is written to other DataNodes. As a result, writing files is slow or fails.

**Solution**

**Step 1** Run the `ulimit -a` command to check the maximum number of file handles set for the involved node. If the value is small, change it to 640000.
Step 2 Run the `vi /etc/security/limits.d/90-nofile.conf` command to edit this file. Set the number of file handles to **64000**. If the file does not exist, create one and modify the file as follows:

![Figure 13-10 Changing the number of file handles](image)

Step 3 Open another terminal. Run the `ulimit -a` command to check whether the modification is successful. If the modification fails, perform the preceding operations again.

Step 4 Restart the DataNode instance on MRS Manager.

--- End

### 13.14 A Client File Fails to Be Closed After Data Writing

**Symptom**

A client file fails to be closed after data is written to the file. A message is displayed indicating that the data block does not have enough replicas.

Client log:

```
20150527/10/6_20150527105000_20150527105500_SR5S14_1432723806338_128_11.pkg.tmp143272380633
8 close hdfs sequence file fail (SequenceFileInfoChannel.java:444)
java.io.IOException: Unable to close file because the last block does not have enough number of replicas.
at org.apache.hadoop.hdfs.DFSOutputStream.completeFile(DFSOutputStream.java:2160)
at org.apache.hadoop.hdfs.DFSOutputStream.close(DFSOutputStream.java:2128)
at org.apache.hadoop.fs.FSDataOutputStream.close(FSDataOutputStream.java:103)
at com.huawei.pai.collect2.stream.SequenceFileInfoChannel.close(SequenceFileInfoChannel.java:433)
at com.huawei.pai.collect2.stream.SequenceFileWriterToolChannel
```
Cause Analysis

1. The HDFS client starts to write blocks.

For example, the HDFS client starts to write /20150527/10_6_20150527105000_20150527105500_SR5S14_1432723806338_128_11.pkg.tmp1432723806338 at 2015-05-27 18:50:24,232. The allocated block is blk_1099105501_25370893:

```
```

2. After the writing is complete, the HDFS client invokes fsync:

```
```

3. The HDFS client invokes close to close the file. After receiving the close request from the client, the NameNode uses the checkFileProgress function to check the completion status of the last block and closes the file only if enough DataNodes report that the last block is complete:

```
2015-05-27 19:00:27,603 | INFO  | IPC Server handler 44 on 25000 | BLOCK* checkFileProgress: blk_1099105501_25370893{blockUCState=COMMITTED, primaryNodeIndex=-1, replicas={[ReplicaUnderConstruction[[DISK]DS-b27b7b7b70-f610-4958-8eba-6decbca2f877:NORMAL|RBW], ReplicaUnderConstruction[[DISK]DS-76bd80e7-ad58-49c6-b2c0-3f9f1ca7f50-NORMAL|RBW]]} has not reached minimal replication 1 | org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkFileProgress(FSNamesystem.java:3197)
2015-05-27 19:00:28,005 | INFO  | IPC Server handler 22 on 25000 | BLOCK* checkFileProgress: blk_1099105501_25370893 has not reached minimal replication 1 | org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkFileProgress(FSNamesystem.java:3197)
2015-05-27 19:00:30,408 | INFO  | IPC Server handler 63 on 25000 | BLOCK* checkFileProgress: blk_1099105501_25370893 has not reached minimal replication 1 | org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkFileProgress(FSNamesystem.java:3197)
2015-05-27 19:00:33,610 | INFO  | IPC Server handler 44 on 25000 | BLOCK* checkFileProgress: blk_1099105501_25370893 has not reached minimal replication 1 | org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkFileProgress(FSNamesystem.java:3197)
```
replicas=[ReplicaUnderConstruction[DISK]DS-ef5fd3c9-5088-4813-ae9a-34a0714ec3a3:NORMAL|RBW], ReplicaUnderConstruction[DISK]DS-f863e30f-ce5b-48cc-9cca-72f64c58adc:NORMAL|RBW])
has not reached minimal replication 1 |
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkFileProgress(FSNamesystem.java:3197)

4. The NameNode prints *CheckFileProgress* multiple times because the HDFS client retries to close the file for several times. The file closing fails because the block status is not complete. The number of retries is determined by the *dfs.client.block.write.locateFollowingBlock.retries* parameter. The default value is 5. Therefore, *CheckFileProgress* is printed six times in the NameNode log.

5. After 0.5 second, the DataNodes report that the block has been successfully written.

6. The block write success notification is delayed because of network bottlenecks or CPU bottlenecks.

7. If close is invoked again or the number of file closing retries increases, a closing success message will be displayed. You are advised to increase the value of *dfs.client.block.write.locateFollowingBlock.retries*. The default parameter value is 5 and retry intervals are 400 ms, 800 ms, 1600 ms, 3200 ms, 6400 ms, and 12800 ms. Therefore, the result of the close function can be returned after a maximum of 25.2 seconds.

**Solution**

**Step 1** Solution:

Set the value of *dfs.client.block.write.locateFollowingBlock.retries* to 6. The retry intervals are 400 ms, 800 ms, 1600 ms, 3200 ms, 6400 ms, and 12800 ms. Therefore, the result of the close function can be returned after a maximum of 50.8 seconds.

----End

**Remarks**

Generally, this fault occurs when the cluster workload is heavy. Adjusting the parameter can only temporarily avoid the fault. You are advised to reduce the cluster workload, for example, do not allocate all CPU resources to MapReduce.

---
13.15 File Fails to Be Uploaded to HDFS Due to File Errors

**Symptom**

The `hadoop dfs -put` command is used to copy local files to HDFS.

After some files are uploaded, an error occurs. The size of the temporary files no long changes on the native NameNode page.

**Cause Analysis**

1. Check the NameNode log `/var/log/Bigdata/hdfs/nn/hadoop-omm-namenode-host name.log`. It is found that the file is being written until a failure occurs.

   ```
   2015-07-13 10:05:07,847 | WARN  | org.apache.hadoop.hdfs.server.namenode.LeaseManager
   $Monitor@36fea922 | DIR* NameSystem.internalReleaseLease: Failed to release lease for file /hive/order/OS_ORDER._8.txt._COPYING_. Committed blocks are waiting to be minimally replicated. Try again later. | FSNamesystem.java:3936
   $Monitor@36fea922 | Cannot release the path /hive/order/OS_ORDER._8.txt._COPYING_ in the lease [Lease. Holder: DFSClient_NONMAPREDUCE_-1872896146_1, pendingcreates: 1] | LeaseManager.java:459
   ```

   org.apache.hadoop.hdfs.protocol.AlreadyBeingCreatedException: DIR*
   NameSystem.internalReleaseLease: Failed to release lease for file /hive/order/OS_ORDER._8.txt._COPYING_. Committed blocks are waiting to be minimally replicated. Try again later.
   at FSNamesystem.internalReleaseLease(FSNamesystem.java:3937)

2. Root cause: The uploaded files are damaged.

3. Verification: The `cp` or `scp` operation fails to be performed for the copied files. Therefore, the files are damaged.

**Solution**

**Step 1** Upload normal files.

----End

13.16 After dfs.blocksize Is Configured on the Interface and Data Is Put, the Original Block Size Keeps Unchanged

**Symptom**

After dfs.blocksize is set to 268435456 on the interface and data is put, the original block size keeps unchanged.

**Cause Analysis**

The `dfs.blocksize` value in the `hdfs-site.xml` file of the client is not changed, and the value prevails.
Solution

Step 1  Ensure that the `dfs.blocksize` value is a multiple of 512.

Step 2  Download a client or modify the client configuration.

Step 3  `dfs.blocksize` is configured on the client and is subject to the client. Otherwise, the value configured on the server prevails.

13.17 Failed to Read Files, and "FileNotFoundException" Is Displayed

Symptom

In MapReduce tasks, all Map tasks are successfully executed, but Reduce tasks fail. The error message "FileNotFoundException...No lease on...File does not exist" is displayed in the logs.

```java
Error: org.apache.hadoop.ipc.RemoteException(java.io.FileNotFoundException): No lease on /user/sparkhive/warehouse/daas/dsp/output/_temporary/1/_temporary/attempt_1479799053892_17075_r_000007_0/part-r-00007 (inode 6501287): File does not exist. Holder DFSClient_attempt_1479799053892_17075_r_000007_0_-1463597952_1 does not have any open files.
at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkLease(FSNamesystem.java:3350)
at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.completeFileInternal(FSNamesystem.java:3442)
at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.completeFile(FSNamesystem.java:3409)
at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.complete(NameNodeRpcServer.java:789)
```

Cause Analysis

"FileNotFoundException...No lease on...File does not exist" indicates that the file is deleted during the operation.

1. Search for the file name in the NameNode audit log of HDFS (`/var/log/Bigdata/audit/hdfs/nn/hdfs-audit-namenode.log` of the active NameNode) to confirm the creation time of the file.

2. Search the NameNode audit logs that are generated within the time range from the file creation to the time of exception occurrence and determine whether the file is deleted or moved to another directory.

3. If the file is not deleted or moved, the parent directory of the file may be deleted or moved. You need to search the upper-layer directory. In this example, the parent directory of the file's parent directory is deleted.

```bash
2017-05-31 02:04:08,286 | INFO  | IPC Server handler 30 on 25000 | allowed=true ugi=appUser@HADOOP.COM (auth:TOKEN) ip=/192.168.1.22 cmd=delete src=/user/sparkhive/warehouse/daas/dsp/output/_temporary dst=null perm=null proto=rpc FSNamesystem.java:8189
```

NOTE

- The preceding log indicates that the `appUser` user of the 192.168.1.22 node deletes `/user/sparkhive/warehouse/daas/dsp/output/_temporary`.
- Run the `zgrep "file name" *.zip` command to search for the contents of the .zip package.
Solution

Step 1 Check the service to find out why the file or the parent directory of the file is deleted.

----End

13.18 Failed to Write Files to HDFS, and "item limit of / is exceeded" Is Displayed

Symptom

The client or upper-layer component logs indicate that a file fails to be written to a directory on HDFS. The error information is as follows:

The directory item limit of /tmp is exceeded: limit=5 items=5.

Cause Analysis

1. The run log file /var/log/Bigdata/hdfs/nn/hadoop-omm-namenode-XXX.log of the client or NameNode contains error information "The directory item limit of /tmp is exceeded:" The error message indicates that the number of files in the /tmp directory exceeds 1048576.

2018-03-14 11:18:21,625 | WARN | IPC Server handler 62 on 25000 | DIR* NameSystem.startFile: /tmp/test.txt The directory item limit of /tmp is exceeded: limit=1048576 items=1048577 | FSNamesystem.java:2334

2. The dfs.namenode.fs-limits.max-directory-items parameter specifies the maximum number of directories or files that are not in recursion relationship in a single directory. The default value is 1048576. The value ranges from 1 to 6400000.

Solution

Step 1 Check whether it is normal that the directory contains more than one million files that are not in recursion relationship. If it is normal, increase the value of the HDFS parameter dfs.namenode.fs-limits.max-directory-items and restart the HDFS NameNode for the modification to take effect.

Step 2 If it is abnormal, delete unnecessary files.

----End

13.19 Adjusting the Log Level of the Shell Client

- Temporary adjustment: After the Shell client window is closed, the log is restored to the default value.
  a. Run the export HADOOP_ROOT_LOGGER command to adjust the log level of the client.
  b. Run the export HADOOP_ROOT_LOGGER=LOG level,console command to adjust the log level of the Shell client.
Run the `export HADOOP_ROOT_LOGGER=DEBUG,console` command to adjust the log level to Debug.

Run the `export HADOOP_ROOT_LOGGER=ERROR,console` command to adjust the log level to Error.

- **Permanent adjustment**
  
  a. Add `export HADOOP_ROOT_LOGGER=\log level,console` to the HDFS client's environment variable configuration file `/opt/client/HDFS/component_env` (replace `/opt/client` with the real-world client path).
  
  b. Run the source `/opt/client/bigdata_env` command.
  
  c. Run the command on the client again.

### 13.20 File Read Fails, and "No common protection layer" Is Displayed

#### Symptom

HDFS fails to be operated on the Shell client or other clients, and the error message "No common protection layer between client and server" is displayed.

Running any `hadoop` command, such as `hadoop fs -ls /`, on a node outside the cluster fails. The bottom-layer error message is displayed stating "No common protection layer between client and server."

```java
org.apache.sqoop.core.SqoopServer.initializeServer(SqoopServer.java:69)
org.apache.sqoop.common.SqoopException: MAPRED_EXEC_0028:Failed to operate HDFS - Failed to get the file /user/loader/etl_dirty_data_dir status
at org.apache.sqoop.job.mr.HDFSClient.fileExist(HDFSClient.java:85)
... at java.lang.Thread.run(Thread.java:745)
Caused by: java.io.IOException: Failed on local exception: java.io.IOException: Couldn't setup connection for loader/hadoop@HADOOP.COM to loader37/10.162.0.37:25000; Host Details : local host is: "loader37/10.162.0.37"; destination host is: "loader37":25000;
   at org.apache.hadoop.net.NetUtils.wrapException(NetUtils.java:776)
... 10 more
Caused by: javax.security.sasl.SaslException: No common protection layer between client and server
   at com.sun.security.sasl.gsskerb.GssKrb5Client.doFinalHandshake(GssKrb5Client.java:251)
... 28 more
Caused by: javax.security.sasl.SaslException: No common protection layer between client and server
   at com.sun.security.sasl.gsskerb.GssKrb5Client.doFinalHandshake(GssKrb5Client.java:251)
... 28 more
```

#### Cause Analysis

1. The RPC protocol is used for data transmission between the client and server of HDFS. The protocol has multiple encryption modes and the `hadoop.rpc.protection` parameter specifies the mode to use.

2. If the value of the `hadoop.rpc.protection` parameter on the client is different from that on the server, the "No common protection layer between client and server" error is reported.
\[ NOTE \]

The hadoop.rpc.protection indicates that data can be transmitted between nodes in any of the following modes:

- **privacy**: Data is transmitted after authentication and encryption. This is the default value. This mode reduces the performance.
- **authentication**: Data is transmitted after authentication without encryption. This mode ensures performance but has security risks.
- **integrity**: Data is transmitted without encryption or authentication. To ensure data security, exercise caution when using this mode.

**Solution**

**Step 1** Download the client again. If the client is an application, update the configuration file in the application.

---

13.21 Failed to Write Files Because the HDFS Directory Quota Is Insufficient

**Symptom**

After quota is set for a directory, writing files to the directory fails. The "The DiskSpace quota of /tmp/tquota2 is exceeded" error message is displayed.

```bash
[omm@189-39-150-115 client]$ hadoop dfs -put switchuser.py /tmp/tquota2
put: The DiskSpace quota of /tmp/tquota2 is exceeded: quota = 157286400 B = 150 MB but diskspace consumed = 402653184 B = 384 MB
```

**Possible Causes**

The remaining space configured for the directory is less than the space required for writing files.

**Cause Analysis**

1. The HDFS supports setting the quota for a specific directory, that is, the maximum space occupied by files in a directory can be set. For example, the following command is used to set a maximum of 150 MB files to be written to the /tmp/tquota directory. (Space = Block size x Number of copies)

   ```bash
   hadoop dfsadmin -setSpaceQuota 150M /tmp/tquota2
   ```

2. Run the following command to check the configured quota for the directory. SPACE_QUOTA is the configured space quota, and REM_SPACE_QUOTA is the remaining space.

   ```bash
   hdfs dfs -count -q -h -v /tmp/tquota2
   ```

**Figure 13-11** Viewing the quota set for a directory
3. Analyze logs. The following log indicates that writing the file requires 384 MB space, but the current space quota is only 150 MB. Therefore, the space is insufficient. Before a file is written, the required remaining space is as follows:

Block size x Number of copies. 128 MB x 3 copies = 384 MB.

```
[omm@189-39-150-115 client]$ hadfs dfs -put switchuser.py /tmp/tquota2
put: The DiskSpace quota of /tmp/tquota2 is exceeded: quota = 157286400 B = 150 MB but diskspace consumed = 402653184 B = 384 MB
```

Solution

**Step 1** Set a proper quota for the directory.

```
hadoop dfsadmin -setSpaceQuota 150G /directory name
```

**Step 2** Run the following command to clear the quota:

```
hdfs dfsadmin -clrSpaceQuota /directory name
```

----End

13.22 Balancing Fails, and "Source and target differ in block-size" Is Displayed

Symptom

When the `distcp` command is executed to copy files across clusters, the message "Source and target differ in block-size." is displayed, indicating that some files fail to be copied. Use `-pb` to preserve block-sizes during copy.

```
at org.apache.hadoop.tools.mapred.RetriableFileCopyCommand.compareCheckSums(RetriableFileCopyCommand.java:214)
```

Possible Causes

This is not a version-related problem. When you run the `distcp` command to copy files, the block size of the source file is not recorded by default. As a result, the verification fails when the block size of the source file is not 128 MB. In this case, you need to add parameter `-pb` to the `distcp` command.

Cause Analysis

1. The block size is set when data is written to HDFS. The default block size is 128 MB. The size of files written by some components or service programs may not be 128 MB, for example, 8 MB.

```
<name>dfs.blocksize</name>
<value>134217728</value>
```
Figure 13-12 Size of files written by some components or service programs

<table>
<thead>
<tr>
<th>Permission</th>
<th>Owner</th>
<th>Group</th>
<th>Size</th>
<th>Last Modified</th>
<th>Replication</th>
<th>Block Size</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>---</td>
<td>bill</td>
<td>hive</td>
<td>15.5 MB</td>
<td>Wed Dec 11 12:44:17</td>
<td>3</td>
<td>8 MB</td>
<td></td>
</tr>
</tbody>
</table>

2. DistCp reads the file from a source cluster and writes it to a destination cluster. By default, the value of dfs.blocksize in the MapReduce task is used as the block size, whose default value is 128 MB.

3. After DistCp finishes writing a file, the system performs verification based on the physical size of the block. Because the block size of the file in the source cluster is different from that of the file in the destination cluster, the splitting sizes are different. As a result, the verification fails.

For example, in the preceding file, there are three blocks \(17.9/8\text{ MB} = 3\) blocks) in the old cluster and one block \(17.9/128\text{ MB} = 1\) block) in the new cluster. Therefore, the verification fails because the physical size of the disk is divided.

Solution

Add parameter `-pb` in the `distcp` command. This parameter is used to reserve the block size when `distcp` is used to ensure that the block size of the new cluster is the same as that of the old cluster.

Figure 13-13 Size of the reserved block during `distcp` command execution

13.23 A File Fails to Be Queried or Deleted, and the File Can Be Viewed in the Parent Directory (Invisible Characters)

Symptom

A file fails to be queried or deleted using the HDFS Shell client. The file can be viewed in the parent directory.

Figure 13-14 List of files in the parent directory
Cause Analysis

The possible cause is that invisible characters are written to the file. You can write the file name to the local text and run the `vi` command to open the file.

```
hdfs dfs -ls parent directory > /tmp/t.txt
vi /tmp/t.txt
```

Run the `set list` command to display invisible characters in the file name. For example, the file name contains `^M`, which is invisible.

![Figure 13-15 Displaying invisible characters](image)

Solution

**Step 1** Run the Shell command to read the file name recorded in the text. Ensure that the following command output contains the full path of the file in HDFS.

```
cat /tmp/t.txt | awk '{print $8}'
```

![Figure 13-16 File path](image)

**Step 2** Run the following command to delete the file:

```
hdfs dfs -rm $(cat /tmp/t.txt | awk '{print $8}')
```

**Step 3** Verify that the file has been deleted.

```
hdfs dfs -ls parent directory
```

----End

13.24 Uneven Data Distribution Due to Non-HDFS Data Residuals

**Symptom**

Data distribution is uneven. A disk is full while other disks have sufficient space.

The data storage directory of HDFS DataNode is set to `/export/data1/dfs--/
/export/data12/dfs`. A large volume of data is stored to `/export/data1/dfs` but data is evenly distributed to other disks.
Cause Analysis

The customer's disk is reinstalled. However, a directory is not thoroughly deleted during disk uninstallation, that is, the added disk is unformatted and historical junk data remains.

Solution

Manually delete data residuals.

13.25 Uneven Data Distribution Due to the Client Installation on the DataNode

Symptom

Data is unevenly distributed on HDFS DataNodes. Disk usage of a node is high or even reaches 100% while disks on other nodes have sufficient idle space.

Cause Analysis

In the HDFS data replica mechanism, the first replica is stored to the local node where the client is stored. As a result, disks of the node run out while disks of other nodes have sufficient idle space.

Solution

Step 1 For the existing data unevenly distributed, run the following command to balance data:

```
/opt/client/HDFS/hadoop/sbin/start-balancer.sh -threshold 10
```

/ opt/client indicates the actual client installation directory.

Step 2 For new data, install the client on the node without DataNode.

---End

13.26 Handling Unbalanced DataNode Disk Usage on Nodes

Symptom

The disk usage of each DataNode on a node is uneven.

The following provides an example:

```
189-39-235-71:~ # df -h
Filesystem  Size  Used  Avail  Use%  Mounted on
/dev/xvda  360G  92G   250G  28%  /
/dev/xvdb  700G  900G   200G  78%  /srv/BigData/hadoop/data1
/dev/xvdc  700G  900G   200G  78%  /srv/BigData/hadoop/data2
/dev/xvdd  700G  900G   200G  78%  /srv/BigData/hadoop/data3
/dev/xvde  700G  900G   200G  78%  /srv/BigData/hadoop/data4
```
Possible Causes

Some disks are faulty and are replaced with new ones. The new disk usage is low. Disks are added. For example, the original four data disks are expanded to five disks.

Cause Analysis

There are two policies for writing data to Block disks on DataNodes: 1. Round Robin (default value) and 2. Preferentially writing data to the disk with the more available space.

Description of the `dfs.datanode.fsdataset.volume.choosing.policy` parameter

Possible values:

- Polling: `org.apache.hadoop.hdfs.server.datanode.fsdataset.RoundRobinVolumeChoosingPolicy`
- Preferentially writing data to the disk with more available space: `org.apache.hadoop.hdfs.server.datanode.fsdataset.AvailableSpaceVolumeChoosingPolicy`

Solution


In this way, the DataNode preferentially selects a node with the most available disk space to store data copies.

**NOTE**

- Data written to the DataNode will be preferentially written to the disk with more available disk space.
- The high usage of some disks can be relieved with the gradual deletion of aging data from the HDFS.

13.27 Locating Common Balance Problems (Updated on July 27, 2019)

Problem 1: Lack of Permission to Execute the balance Task (Access denied).

Problem details: After the `start-balancer.sh` command is executed, the "hadoop-root-balancer-host name.out" log displays "Access denied for user test1. Superuser privilege is required."

```
cat /opt/client/HDFS/hadoop/logs/hadoop-root-balancer-host2.out
```
INFO: Watching file:/opt/client/HDFS/hadoop/etc/hadoop/log4j.properties for changes with interval: 60000


Superuser privilege is required
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkSuperuserPrivilege(FSPermissionChecker.java:122)
at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkSuperuserPrivilege(FSNamesystem.java:5916)

Cause analysis:
The administrator account is required for executing the balance task.

Solution
- Secure version
  Perform authentication for user hdfs (default password: Hdfs@123) or other users in the supergroup group and then execute the balance task.
- General version
  Run the su - hdfs command on the client before running the balance command on HDFS.

Problem 2: The balance command fails to be executed, and the /system/balancer.id file is abnormal.

Problem details:
A user starts a balance process on the HDFS client. After the process is stopped unexpectedly, the user performs the balance operation again. The operation fails.


Cause analysis:
Generally, after the balance operation is complete in HDFS, the /system/balancer.id file is automatically released and the balance operation can be performed again.

In the preceding scenario, the first balance operation is stopped abnormally. Therefore, when the balance operation is performed for the second time, the /system/balancer.id file still exists. As a result, the append /system/balancer.id operation is triggered and the balance operation fails.

Solution
Method 1: After the hard lease period exceeds one hour, release the lease on the original client and perform the balance operation again.

Method 2: Delete the /system/balancer.id file from HDFS and perform the balance operation again.
13.28 HDFS Displays Insufficient Disk Space But 10% Disk Space Remains

Symptom

1. The alarm "ALM-14001 HDFS Disk Usage Exceeds the Threshold" is reported.
2. On the HDFS page, high disk space usage is displayed.

Cause Analysis

The `dfs.datanode.du.reserved.percentage` parameter is set in HDFS, indicating the percentage of the reserved space of each disk to the total disk space. The DataNode reserves space you set for NodeManager running and computing of other components, for example, Yarn, or for upgrades.

As 10% disk space is reserved, the HDFS DataNode regards that there is no available disk space when the disk usage reaches 90%.

Solution

Step 1 Expand the HDFS DataNode disk capacity when its usage reaches 80%. For details about how to expand the disk capacity, see Adding a New Disk to an MRS Cluster.

Step 2 If the disk capacity cannot be expanded in time, delete useless data in HDFS to release disk space.

----End

13.29 An Error Is Reported When the HDFS Client Is Installed on the Core Node in a Common Cluster

Issue

When a user is created on a Core node in a common cluster and the client is used for installation, a misleading error is reported.

Symptom

In a common cluster, the following error message is displayed when a user is created on the Core node to install the client:

```
2020-03-14 19:16:17,166 WARN shortcircuit.DomainSocketFactory: error creating DomainSocket
java.net.ConnectException: connect(2) error: Permission denied when trying to connect to '/var/run/MRS-HDFS/dn_socket'
    at org.apache.hadoop.net.unix.DomainSocket.connect0(Native Method)
    at org.apache.hadoop.net.unix.DomainSocket.connect(DomainSocket.java:256)
    at org.apache.hadoop.hdfs.shortcircuit.DomainSocketFactory.createSocket(DomainSocketFactory.java:168)
    at org.apache.hadoop.hdfs.client.impl.BlockReaderFactory.nextDomainPeer(BlockReaderFactory.java:799)
```
Cause Analysis

A user runs the `useradd` command to create a user. The default user group of the user does not contain the `ficommmon` user group. As a result, the preceding error is reported when the `get` command of HDFS is executed.

Procedure

Run the `usermod -a -G ficommmon username` command to add the user to the `ficommmon` user group.
14.1 Plenty of Jobs Are Found After Yarn Is Started

Issue
After the customer creates an MRS cluster and starts Yarn, plenty of jobs occupying resources are found.

Symptom
After the customer creates an MRS cluster and starts Yarn, plenty of jobs occupying resources are found.

Cause Analysis
- It is suspected that there are hacker attacks.
- Set the Any protocol in the inbound direction of the SG to the 0.0.0.0/0.
Procedure

Step 1  Log in to the MRS management console. On the Active Clusters page, click the cluster name. The cluster details page is displayed.

Step 2  Click Manage next to Cluster Manager. The Access MRS Manager page is displayed.

Step 3  Click Manage Security Group Rule to check the security group rule configuration.

Step 4  Check whether the source address of the Any protocol in the inbound direction is 0.0.0.0/0.

Step 5  If it is 0.0.0.0/0, change the remote end of the Any protocol in the inbound direction to a specified IP address. If it is not 0.0.0.0/0, there is no need to change the value.

Step 6  After the value is changed successfully, restart the cluster VM.

Summary and Suggestions

Disable the Any protocol in the inbound direction, or specify the remote end of the Any protocol in the inbound direction as the specified IP address.

Related Information

For details, see Security Configuration Suggestions for Clusters with Kerberos Authentication Disabled.
14.2 Error Message "GC overhead" Is Displayed on the Client During Task Submission by Using the Hadoop Jar Command

Symptom

When a user submits a task on the client, the client returns a memory overflow error.

Cause Analysis

According to the error stack, the memory overflows when the HDFS files are read during task submission. Generally, the memory is insufficient because the task needs to read a large number of small files.

Solution

**Step 1** Check whether multiple HDFS files need to be read for the started MapReduce tasks. If yes, reduce the file quantity by combining the small-sized files in advance or using `combineInputFormat`.

**Step 2** Increase the memory used for Hadoop command execution. Specifically, change the value of `HADOOP_CLIENT_OPTS` (default value: `512m`) to a larger one, for example, `5120m`. The memory is set in the `/opt/client/HDFS/hadoop/etc/hadoop/hadoop-env.sh` path on the client. `/opt/client` refers to the actual client path.
14.3 Disk Space Is Used Up Due to Oversized Aggregated Logs of Yarn

**Issue**

The disk usage of the cluster is high.

**Symptom**

- On the host management page of MRS Manager, the disk usage of the Core nodes is too high.
- Only a few tasks are running on the Yarn web UI.
- Log in to the Master node of the cluster and run the `hdfs dfs -du -h /` command. The following files occupy a large amount of disk space:

<table>
<thead>
<tr>
<th>File Path</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0153</td>
<td>22.5 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0154</td>
<td>18.4 M</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0155</td>
<td>23.4 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0156</td>
<td>23.5 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0157</td>
<td>23.7 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0158</td>
<td>23.7 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0159</td>
<td>22.5 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0160</td>
<td>23.7 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0161</td>
<td>23.7 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0162</td>
<td>23.7 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0163</td>
<td>22.5 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0164</td>
<td>23.7 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0165</td>
<td>23.7 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0166</td>
<td>23.7 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0167</td>
<td>23.7 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0168</td>
<td>23.7 G</td>
</tr>
<tr>
<td>/tmp/logs/root/logs/application_1589278244986_0169</td>
<td>23.7 G</td>
</tr>
</tbody>
</table>

- The log aggregation configuration of the Yarn service is as follows:

```
* yarn.log-aggregation.retain-check-interval-seconds 86400
* yarn.log-aggregation.retain-seconds 129600
```

**Cause Analysis**

The customer submits tasks too frequently, and the time for deleting aggregated log files is set to 1296000, that is, aggregated logs are retained for 15 days. As a
result, aggregated logs cannot be released within a short period of time, exhausting disk space.

Procedure

**Step 1** Logging in to MRS Manager

**Step 2** Choose Services > Yarn > Service Configuration. Set Type to All.

**Step 3** Search for the `yarn.log-aggregation.retain-seconds` parameter and decrease the value of `yarn.log-aggregation.retain-seconds` based on the site requirements, for example, to 259200. That is, the aggregated logs of Yarn are retained for three days and the disk space is automatically released after the retention period expires.

**Step 4** Click Save Configuration and deselect Restart the affected services or instances.

**Step 5** Restart the Yarn service during off-peak hours. Restarting the Yarn service will interrupt upper-layer services and affect cluster management, maintenance, and services.

1. Log in to MRS Manager.
2. Choose Services, locate the row that contains Yarn, and click Restart in the Operation column.

----End
15.1 DBServer Instance Status Abnormal

Symptom

A DBServer instance always in the Concerning state.

Figure 15-1 DBServer instance status

Cause Analysis

The permission for files or directories in the data directory is incorrect. GaussDB requires that the file permission be at least 600 and directory permission be at least 700.

Figure 15-2 Directory permission list

```
...ll
total 4
drwx------ 19 omm wheel 4096 Dec 14 10:15 data
```
Figure 15-3 File permission list

```
omen@192.168-234-1/b/srv/BigData/dbdata_service/data> ll
total 128
drwx------ 6 omm wheel 4096 Dec 9 15:47 base
-rw------- 1 omm wheel 922 Dec 9 15:34 dblink.conf
-rw------- 1 omm wheel 16 Dec 14 10:15 gaussdb.state
drwx------ 2 omm wheel 4096 Dec 14 10:17 global
drwx------ 2 omm wheel 4096 Dec 11 00:00 pg_audit
drwx------ 2 omm wheel 4096 Dec 14 10:15 pg_blackbox
drwx------ 2 omm wheel 4096 Dec 9 15:34 pg_clog
drwx------ 2 omm wheel 4096 Dec 14 10:15 pg_confrole_backup
d-rw------- 1 omm wheel 1024 Dec 9 15:34 pg_ctl.lock
-rw------- 1 omm wheel 4245 Dec 9 15:47 pg_hba.conf
-rw------- 1 omm wheel 1024 Dec 9 15:47 pg_hba.conf.lock
-rw------- 1 omm wheel 1636 Dec 9 15:34 pg_ident.conf
drwx------ 2 omm wheel 4096 Dec 9 15:38 pg_log
drwx------ 4 omm wheel 4096 Dec 9 15:34 pg_multixact
drwx------ 2 omm wheel 4096 Dec 14 10:15 pg_notify
drwx------ 2 omm wheel 4096 Dec 9 15:34 pg_serial
drwx------ 2 omm wheel 4096 Dec 9 15:34 pg_snapshots
drwx------ 2 omm wheel 4096 Dec 14 11:56 pg_stat_tmp
drwx------ 2 omm wheel 4096 Dec 9 15:34 pg_subtrans
drwx------ 2 omm wheel 4096 Dec 9 15:34 pg_tblspc
drwx------ 2 omm wheel 4096 Dec 9 15:34 pg_twophase
d-rw------- 1 omm wheel 4 Dec 9 15:34 pg_VERSION
drwx------ 2 omm wheel 4096 Dec 9 15:34 pg_wal

```

Solution

Step 1  Modify the permissions on the files and directories based on the permission list in Figure 15-2 and Figure 15-3.

Step 2  Restart the DBServer instance.

----End

15.2 DBServer Instance Is Always in the Concerning State

Symptom

A DBServer node is always in the Concerning state. The node status cannot be recovered even after a restart.

Figure 15-4 DBServer instance status
Cause Analysis

1. The DBService monitors the /opt/Bigdata/MRS_2.0.0/install/dbservice/ha/module/harm/plugin/script/gsDB/.startGS.fail file.
2. If the value in the file is greater than 3, the startup fails. The NodeAgent keeps trying to restart the instance. In this case, the startup still fails and the value increases by 1 each time the startup fails.

Solution

Step 1 Log in to MRS Manager.
Step 2 Stop the DBServer instance.
Step 3 Log in to the node where the DBServer instance is abnormal as user omm.
Step 4 Change the value in the /opt/Bigdata/MRS_2.0.0/install/dbservice/ha/module/harm/plugin/script/gsDB/.startGS.fail file to 0.
Step 5 Start the DBServer instance.

15.3 Default Port Number 20050 or 20051 Is Occupied

Symptom

DBService restart fails, and information indicating that port 20050 or 20051 is occupied is displayed in the printed fault log.

Cause Analysis

1. The default port 20050 or 20051 used by DBService is occupied by another process.
2. The DBService process is not stopped, and the port used by DBService is not released.

Solution

This solution uses port 20051 as an example. The solution to the problem that port 20050 is occupied is similar.

Step 1 Use PuTTY to log in to the node where the error is reported as user root, and run the netstat -nap | grep 20051 command to check the process that occupies port 20051.
Step 2 Run the kill command to forcibly stop the process that uses port 20051.
Step 3 About 2 minutes later, run the netstat -nap | grep 20051 command again to check whether any process uses the port.
Step 4 Check the service to which the process belongs and change the port for the service.
Step 5  Run the `find . -name "*20051*"` command in the `/tmp` and `/var/run/FusionInsight-DBService` directories, and delete all files found.

Step 6  Log in to MRS Manager, click Services, and restart DBService.

---End

15.4 DBServer Instance Is Always in the Concerning State Because the /tmp Directory Permission Is Incorrect

Symptom

A DBServer node is always in the Concerning state. The status cannot be recovered even after a restart.

Figure 15-5 DBServer instance status

Cause Analysis

1. Check `/var/log/Bigdata/dbservice/healthCheck/dbservice_processCheck.log`. It is found that GaussDB is abnormal.

Figure 15-6 GaussDB exception

2. The check result shows that the permission on the /tmp directory is incorrect.
**Solution**

**Step 1** Run the following command to modify the `/tmp` permission:

```
chmod 1777 /tmp
```

**Step 2** Wait until the instance status recovers.

----End

### 15.5 DBService Backup Failure

**Symptom**

```
ls /srv/BigData/LocalBackup/default_20190720222358/ -rlth
```

No DBService backup file exists in the backup file path.

**Cause Analysis**

- Check the backup log of DBService in `/var/log/Bigdata/dbservice/scriptlog/backup.log`. It is found that the backup is successful but fails to be uploaded to the OMS node.
The failure is caused by the SSH failure.

Solution

Step 1 If the network is faulty, contact network engineers.

Step 2 Perform backup operations again after the network fault is rectified.

----End

15.6 DBService Is Normal But Components Cannot Connect to It

Symptom

Upper-layer components fail to connect to DBService. The DBService component and two instances are normal.

Figure 15-9 DBService status

Cause Analysis

1. The upper-layer component is DBService connected through dbservice.floatip.

2. Run the `netstat -anp | grep 20051` command on the node where DBServer resides. It is found that the Gauss process of DBService is not bound to the floating IP address during startup, and only the local IP address 127.0.0.1 is listened.

Solution

Step 1 Log in to MRS Manager.

Step 2 Choose Services > DBService and choose More > Restart Service to restart DBService.
Step 3 Run the `netstat -anp | grep 20051` command on the active DBServer node to check whether `dbservice.floatip` is bound.

----End

15.7 DBServer Startup Failure

**Symptom**

DBService components fail to be started and restarts also fail. The instance keeps in the **Concerning** state.

**Figure 15-10 DBService status**

<table>
<thead>
<tr>
<th>Box</th>
<th>Host Name</th>
<th>DBIP Address</th>
<th>Business IP Address</th>
<th>Back</th>
<th>Operating Status</th>
<th>Health Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBServer Unsync</td>
<td>node-master1225</td>
<td>162.168.0.1</td>
<td>162.168.13</td>
<td></td>
<td>Started</td>
<td></td>
</tr>
<tr>
<td>DBServer (standby)</td>
<td>node-master1223</td>
<td>162.168.0.2</td>
<td>162.168.13</td>
<td></td>
<td>Started</td>
<td></td>
</tr>
</tbody>
</table>

**Cause Analysis**

1. Check the DBService logs in `/var/log/Bigdata/dbservice/DB/gs_ctl-current.log`. The following error message is displayed.

```
Jan 21 10:48:48 hadoop122 Bigdata/dbservice/DB> ll /usr/libexec
total 100
lslr-xr-x 2 root  root  4096 Aug  6 2016 bin
lslr-xr-x 2 root  root  4096 Aug  5 2016 root
lslr-xr-x 3 root  root  5800 Sep  29 11:30 dev
lslr-xr-x 3 root  root  0 Sep  20 11:10 econdfs
lslr-xr-x 7 root  root  4096 Sep 22 02:40 ece
lslr-xr-x 3 root  root  0 Sep 11 00:25 fack_corrected
lslr-xr-x 9 root  root  4096 Sep 18 14:39 home
lslr-xr-x 12 root  root  4096 Sep 14 2016 lib
lslr-xr-x 0 root  root  12108 Sep 14 2016 lib64
lslr-xr-x 2 root  root  16384 Aug  7 2016 lost+found
lslr-xr-x 2 root  root  4096 May  5 2016 media
lslr-xr-x 2 root  root  4096 May  5 2016 misc
lslr-xr-x 19 root  root  0 Jun 30 10:04 opt
lslr-xr-x 424 root  root  0 Sep 20 19:13 proc
lslr-xr-x 5 root  root  4096 Sep 23 10:11 root
lslr-xr-x 4 root  root  4096 Aug  7 2016 rrdtool
lslr-xr-x 3 root  root  12188 Sep 14 2016 sbin
lslr-xr-x 2 root  root  4096 May  5 2016 selinux
lslr-xr-x 10 root  root  4096 Nov 15 2016 svc
lslr-xr-x 12 root  root  0 Sep 20 11:19 sys
lslr-xr-x 3 root  root  1 Aug  7 2016 target
lslr-xr-x 6 root  root  4096 Nov 23 18:12 tmp
lslr-xr-x 15 root  root  4096 Apr  22 2014 var
```

2. It is found that the `/tmp` permission is incorrect. The correct value should be 777.
Solution

Step 1 Modify the /tmp permission by changing the value to 777.

Step 2 Restart DBService.

----End

15.8 DBService Backup Failure Due to Unreachable Floating IP Address

Symptom

The default DBService backup fails, but backups of NameNode, LdapServer, and OMS are successful.

Cause Analysis

1. Check the error information on the DBService backup page:
   Clear temporary files at backup checkpoint DBService_test_DBService_DBService_20180326155921 that failed last time.
   Temporary files at backup checkpoint DBService_test_DBService_DBService_20180326155921 that failed last time are cleared successfully.

   Start executing the backup task.
   Check the backup available disk space.
   Backup installation processed for configuration DBService.
   Clear temporary files at backup checkpoint DBService_test_DBService_DBService_20180326155921 that failed last time.
   Temporary files at backup checkpoint DBService_test_DBService_DBService_20180326155921 that failed last time are cleared successfully.

   Start backup succeeded for checkpoint DBService_test_DBService_DBService_20180326155921.
   Backup is being performed for checkpoint DBService_test_DBService_DBService_20180326155921.
   Backup execution failed. Task id: 2
   Detail: DBService backup task failed, please view details in logs.
   Temporary files are cleared successfully after backup checkpoint DBService_test_DBService_DBService_20180326155921.
   checkpoint DBService_test_DBService_DBService_20180326155921 is deleted successfully after backup failure.
   Failed to backup configuration DBService.

2. Check the /var/log/Bigdata/dbservice/scriptlog/backup.log file. It is found that the log printing stops and no related backup information is found.

3. Check the /var/log/Bigdata/controller/backupplugin.log file on the active OMS node. The following error information is found:
   result error is ssh:connect to host 172.16.4.200 port 22: Connection refused (172.16.4.200 is the floating IP address of DBService)
   DBService backup failed.

Solution

Step 1 Log in to the active DBService node (the Master node bound with the DBService floating IP address).
Step 2 Add the DBService floating IP address to `ListenAdress` or comment out `ListenAdress` in the `/etc/ssh/sshd_config` file.

Step 3 Run the following command to restart the SSHD service:
```
service sshd restart
```

Step 4 Check whether the next DBService backup is successful.

----End

15.9 Startup Failure Due to Lost DBService Configuration File

Symptom
The nodes are powered off unexpectedly, and the standby DBService node fails to be restarted.

Cause Analysis
1. The `var/log/Bigdata/dbservice/DB/gaussdb.log` file is viewed, which contains no information.

2. The `var/log/Bigdata/dbservice/scriptlog/preStartDBService.log` file is viewed. This file contains the following information, indicating that the configuration information is lost:
The program "gaussdb" was found by "/opt/Bigdata/MRS_2.0.1/install/dbservice/gaussdb/bin/gs_guc"
But not was not the same version as gs_guc.
Check your installation.
3. The configuration file in the `/srv/BigData/dbdata_service/data` directory on the active DBServer node is compared with the configuration file in the `/srv/BigData/dbdata_service/data` directory on the standby DBServer node, which shows major difference.

```
ll
```

```
```

Solution

Step 1 Copy the content in the `/srv/BigData/dbdata_service/data` directory on the active node to the standby node and ensure that the file permission and owner group are the same as those on the active node.

Step 2 Modify configuration in `postgresql.conf`, set the value of `localhost` to the local node, and set `remotehost` to the peer node.
Step 3  Log in to MRS Manager and restart the standby DBServer node.
16.1 "IllegalConfigurationException: Error while parsing YAML configuration file: "security.kerberos.login.keytab" Is Displayed When a Command Is Executed on an Installed Client

Symptom

After the client is successfully installed, an error message "IllegalConfigurationException: Error while parsing YAML configuration file:"security.kerberos.login.keytab" is displayed when the command (for example, yarn-session.sh) on the client is executed.

Causes

1. The following authentication modes are available for Flink.
   - Kerberos authentication: Flink Yarn client, Yarn ResourceManager, JobManager, HDFS, TaskManager, Kafka, and ZooKeeper
   - Internal authentication mechanism of Yarn: The internal authentication used between YarnResourceManager and Application Master (AM).
2. If a security cluster is required, the Kerberos authentication and security cookie authentication are mandatory. As shown in the logs, it is found that the `security.kerberos.login.keytab` setting in the configuration file is incorrect and the security configuration is not performed.

**Solution**

**Step 1**  Download the keytab file from MRS and save it in a folder on a host where the Flink client resides.

**Step 2**  Configure following parameters in the `flink-conf.yaml` file:

1. Keytab path
   ```yaml
   security.kerberos.login.keytab: /home/flinkuser/keytab/abc222.keytab
   ```

   **NOTE**
   - `/home/flinkuser/keytab/abc222.keytab` indicates the user directory, which is the directory saves the keytab file in **Step 1**.
   - Ensure that the client user has the permission on the corresponding directory.

2. Principal name
   ```yaml
   security.kerberos.login.principal: abc222
   ```

3. In HA mode, if Zookeeper is configured, the ZooKeeper Kerberos authentication configuration items must be configured as follows:
   ```yaml
   zookeeper.sasl.disable: false
   security.kerberos.login.contexts: Client
   ```

4. If Kerberos authentication is required between the Kafka client and Kafka broker, configure it as follows:
   ```yaml
   security.kerberos.login.contexts: Client,KafkaClient
   ```

----End

### 16.2 "IllegalConfigurationException: Error while parsing YAML configuration file" Is Displayed When a Command Is Executed After Configurations of the Installed Client Are Changed

**Symptom**

After the client is successfully installed, an error message "IllegalConfigurationException: Error while parsing YAML configuration file: 81: "security.kerberos.login.principal:pippo " is displayed when the command (for example, `yarn-session.sh`) on the client is executed.

```
[root@8-5-131-10 bin]# yarn-session.sh
at org.apache.flink.configuration.GlobalConfiguration.loadConfiguration (GlobalConfiguration.java:112)
at org.apache.flink.configuration.GlobalConfiguration.loadConfiguration (GlobalConfiguration.java:79)
```
Cause Analysis

The `security.kerberos.login.principal:pippo` item in the `flink-conf.yaml` configuration file was faulty.

Solution

Modify the configuration in the `flink-conf.yaml` file.

Note: The configuration item name and value must be separated by a space.

16.3 The `yarn-session.sh` Command Fails to Be Executed When the Flink Cluster Is Created

Symptom

During the creation of the Flink cluster, an error message is displayed after the `yarn-session.sh` command execution is suspended.

Possible Causes

The SSL communication encryption is enabled for Flink, but no correct SSL certificate is configured.
Solution

Method 1:
Run the following command to disable the Flink SSL communication encryption, and modify the client configuration file `conf/flink-conf.yaml`.

```
security.ssl.internal.enabled: false
```

Method 2:
Enable the Flink SSL communication encryption and retain the default value of `security.ssl.internal.enabled`. Configure the SSL as follows:

- If the KeyStore or TrustStore file is a relative path, and the Flink client directory where the command is executed can directly access this relative path.

  ```
  security.ssl.internal.keystore: ssl/flink.keystore
  security.ssl.internal.truststore: ssl/flink.truststore
  ```

  Add `-t` option to the CLI `yarn-session.sh` command of Flink to transmit the KeyStore and TrustStore files to each execution node. For example,

  `yarn-session.sh -t ssl/ -n 2`

- If the keystore or truststore file path is an absolute path, the keystore or truststore files must exist in the absolute path on Flink Client and all nodes.

  ```
  security.ssl.internal.keystore: /opt/client/Flink/flink/conf/flink.keystore
  security.ssl.internal.truststore: /opt/client/Flink/flink/conf/flink.truststore
  ```

16.4 Failed to Create a Cluster by Executing the yarn-session Command When a Different User Is Used

Symptom

Two users `testuser` and `bdpuser` with the same rights are used to create the Flink cluster.

When user `testuser` is used to create a Flink cluster, no error message is displayed. While user `bdpuser` is used to create a Flink cluster, an error message is displayed during the `yarn-session.sh` command execution.

```
2019-01-02 14:28:09,098 | ERROR | [main] | Ensure path threw exception |
org.apache.flink.shaded.zookeeper.org.apache.zookeeper.KeeperException$NoAuthException: KeeperErrorCode = NoAuth for /flink/application_1545397824912_0022
```

Possible Causes

The HA configuration item is not modified. In the Flink configuration file, the default value of `high-availability.zookeeper.client.acl` is `creator`, indicating that only the creator has the access permission. A new user cannot access the directory on ZooKeeper. As a result, the `yarn-session.sh` command execution fails.

Solution

**Step 1** Modify the value of `high-availability.zookeeper.path.root` in the `conf/flink-conf.yaml` file. For example, run the following command:
Step 2  Submit the tasks again.

----End

16.5 Flink Service Program Fails to Read Files on the NFS Disk

Issue

The Flink service program cannot read files on the NFS disk mounted to the cluster node.

Symptom

The Flink service program developed by a user needs to read the user-defined configuration file. The configuration file is stored on the NFS disk. The NFS disk is mounted to the cluster node and can be accessed by all nodes in the cluster. After the user submits the Flink program, the service code cannot access the user-defined configuration file. As a result, the service program fails to be started.

Cause Analysis

The root cause is that the permission on the root directory of the NFS disk is insufficient. As a result, the Flink program cannot access the directory after being started.

Flink tasks of MRS are running on Yarn. If the cluster is a common cluster, the user who runs the tasks on Yarn is yarn_user. If the user-defined configuration file is used after the tasks are started, yarn_user must be allowed to access the file and the parent directory of the file (parent directory of the file on the NFS, not the soft link on the cluster node). Otherwise, the program cannot obtain the file content. If the cluster is a cluster with Kerberos authentication enabled, the file permission must allow the user who submits the program to access the file.

Procedure

Step 1  Log in to the Master node in the cluster as user root.

Step 2  Run the following command to check the permission on the parent directory of the user-defined configuration file:

```
ll <Parent directory of the file path>
```

Step 3  Go to the directory of the file to be accessed on the NFS disk and change the permission of the parent directory of the user-defined configuration file to 755.

```
chmod 755 -R /<Path of the parent directory of the file>
```

Step 4  Check whether the Core or Task node can access the configuration file.

1.  Log in to the Core or Task node as the root user.

   If Kerberos authentication is enabled for the current cluster, log in to the Core node as user root.

```
2. Run `su - yarn_user` to switch to user `yarn_user`.
   If Kerberos authentication is enabled for the current cluster, run the `su - User who submits the job` command to switch the user.

3. Run the following command to check the user permission. The file path must be the absolute path of the file.
   ```
   ll <File path>
   ```

---End

**Summary and Suggestions**

When a user-defined configuration file needs to be accessed in the submitted task, especially when the NFS disk is mounted, you need to check whether the permission of the parent directory of the file is correct in addition to the file permission. When an NFS disk is mounted to an MRS cluster node, a soft link is created to the NFS directory. In this case, you need to check whether the directory permission on the NFS is correct.

**16.6 New Topic 6**
17 Using Impala

17.1 Failed to Connect to impala-shell

Issue

A user fails to connect to impala-shell.

Symptom

After a user modifies the configuration of any component on the component management page and restarts the service, the connection to impala-shell fails, and the error message "no such file/directory" is displayed.

```
/home/user@node:/etc$ impala-shell -i 102.168.0.73
shell-init: error retrieving current directory: getcwd: cannot access parent directories: No such file or directory
```

Cause Analysis

After the service configuration is modified and the service is restarted, some directory structures of the service, such as the etc directory, are deleted and recreated. If the directory is etc or its subdirectory before the service is restarted, some system variables or parameters cannot be found when impala-shell is
executed in the original directory because the directory is recreated after the service is restarted. As a result, impala-shell fails to be connected.

**Procedure**

Switch to any existing directory and reconnect to impala-shell.
18.1 During sql-standard-with-group Configuration, a Schema Fails to Be Created and the Error Message "Access Denied" Is Displayed

Issue
A schema fails to be created during sql-standard-with-group configuration and the error message "Access Denied" is displayed.

Symptom
```
CREATE SCHEMA hive.sf2 WITH (location = 'obs://obs-zy1234/sf2');Query 20200224_031203_00002_g6gzy failed: Access Denied: Cannot create schema sf2
```

Cause Analysis
To create a schema in Presto, you must have the administrator permission of Hive.

Procedure
- Method 1:
  a. Log in to MRS Manager and choose System > Manage User.
  b. Locate the row that contains the target user, and click Modify in the Operation column.
  c. Click Select and Add Role to assign the System_administrator permission to the user.
  d. Click OK.
- Method 2:
  a. Log in to MRS Manager and choose System > Manage Role.
  b. Click Create Role and set the following parameters:
     - Enter a role name, for example, hive_admin.
Set Permission to Hive and select Hive Admin Privilege.
c. Click OK to save the role.
d. Choose System > Manage User.
e. Locate the row that contains the target user, and click Modify in the Operation column.
f. Click Select and Add Role to add the newly created hive_admin permission to the user.
g. Click OK.
19.1 Error Message "Does not contain a valid host:port authority" Is Reported When Alluxio Is in HA Mode

Issue

Error message "Does not contain a valid host:port authority" is reported for Alluxio in HA mode in a security cluster.

Symptom

Error message "Does not contain a valid host:port authority" is reported for Alluxio in HA mode in a security cluster.

Cause Analysis

org.apache.hadoop.security.SecurityUtil.buildDTServiceName does not support multiple alluxiomaster addresses in the URI.

Procedure

Use `alluxio:///` or `alluxio://<IP address or host name of the active AlluxioMaster>:19998/` for access.
20.1 Accessing ZooKeeper from an MRS Cluster

Issue

Accessing ZooKeeper from an MRS cluster

Symptom

The customer uses zkcli.sh to access ZooKeeper on the MRS Master node, but an error is reported.

Cause Analysis

The command used by the customer is incorrect. As a result, an error is reported.

Procedure

Step 1 Obtain the ZooKeeper IP address. For details, see How Do I Obtain the ZooKeeper Address?

Step 2 Log in to the Master node as user root.

Step 3 Run the following command to initialize environment variables:

source /opt/client/bigdata_env

Step 4 Run the zkcli.sh -server "IP address of the node where ZooKeeper is located: 2181" command to connect to ZooKeeper of the MRS cluster.

The IP address of the node where ZooKeeper is located is the one queried in Step 1. Use commas (,) to separate multiple IP addresses.

Step 5 Run common commands such as ls / to view ZooKeeper information.

----End
21 Accessing OBS

21.1 When Using the MRS Multi-user Access to OBS Function, a User Does Not Have the Permission to Access the /tmp Directory

Issue

When the MRS multi-user access to OBS function is used to execute jobs such as Spark, Hive, and Presto jobs, an error message is displayed, indicating that the user does not have the permission to access the /tmp directory.

Symptom

When the MRS multi-user access to OBS function is used to execute jobs such as Spark, Hive, and Presto jobs, an error message is displayed, indicating that the user does not have the permission to access the /tmp directory.

Cause Analysis

A temporary directory exists during job execution. The user who submits the job does not have permission on the temporary directory.

Procedure

Step 1 On the Dashboard tab page of the cluster, query and record the name of the agency bound to the cluster.

Step 2 Log in to the IAM console.

Step 3 Choose Permissions. On the displayed page, click Create Custom Policy.
- **Policy Name**: Enter a policy name.
- **Scope**: Select Global services.
- **Policy View**: Select Visual editor.
- **Policy Content**: 
a. **Allow**: Select **Allow**.
b. **Select service**: Select **Object Storage Service (OBS)**.
c. **Select action**: Select **WriteOnly**, **ReadOnly**, and **ListOnly**.
d. **Specific resources**:
   i. Set **object** to **Specify resource path**, click **Add resource path**, and enter `obs_bucket_name/tmp/` and `obs_bucket_name/tmp/*` in **Path**. The `/tmp` directory is used as an example. If you need to add permissions for other directories, perform the following steps to add the directories and resource paths of all objects in the directories.
   ii. Set **bucket** to **Specify resource path**, click **Add resource path**, and enter `obs_bucket_name` in **Path**.

Replace `obs_bucket-name` with the actual OBS bucket name. If the bucket type is Parallel File System, you need to add the `obs_bucket_name/tmp/` path. If the bucket type is Object Storage, you do not need to add the path.
e. (Optional) Request condition, which does not need to be added currently.

**Figure 21-1** Custom policy

---

**Step 4** Click **OK**.

**Step 5** Select **Agency** and click **Assign Permissions** in the **Operation** column of the agency queried in **Step 1**.

**Step 6** Query and select the created policy in **Step 3**.

**Step 7** Click **OK**.

----End

### 21.2 When the Hadoop Client Is Used to Delete Data from OBS, It Does Not Have the Permission for the .Trash Directory

**Issue**

When a user uses the Hadoop client to delete data from OBS, an error message is displayed indicating that the user does not have the permission on the **.Trash** directory.
**Symptom**

After the `hadoop fs -rm obs://<obs_path>` command is executed, the following error information is displayed:

```
```

**Cause Analysis**

When deleting a file, Hadoop moves the file to the `.Trash` directory. If the user does not have the permission on the directory, error 403 is reported.

**Procedure**

Solution 1:
Run the `hadoop fs -rm -skipTrash` command to delete the file.

Solution 2:
Add the permission to access the `.Trash` directory to the agency corresponding to the cluster.

**Step 1** On the Dashboard tab page of the cluster, query and record the name of the agency bound to the cluster.

**Step 2** Log in to the IAM console.

**Step 3** Choose Permissions > Create Custom Policy.

- **Policy Name**: Enter a policy name.
- **Scope**: Select Global services.
- **Policy View**: Select Visual editor.
- **Policy Content**:
  - **Allow**: Select Allow.
  - **Select service**: Select Object Storage Service (OBS).
  - Select all operation permissions.
  - **Specific resources**:
    - Set object to Specify resource path, click Add resource path, and enter the `.Trash` directory, for example, `obs_bucket_name/user/root/.Trash/*` in Path.
    - Set bucket to Specify resource path, click Add resource path, and enter `obs_bucket_name` in Path.

Replace `obs_bucket-name` with the actual OBS bucket name.

**Figure 21-2 Custom policy**
Step 4  Click OK.

Step 5  Select Agency and click Assign Permissions in the Operation column of the agency queried in Step 1.

Step 6  Query and select the created policy in Step 3.

Step 7  Click OK.

Step 8  Run the hadoop fs -rm obs://<obs_path> command again.

----End