MapReduce Service (MRS)
3.3.1-LTS
O&M Guide

Issue 01

Date 2024-04-30

HUAWEI

HUAWEI CLOUD COMPUTING TECHNOLOGIES CO., LTD.

Trademarks and Permissions


and other Huawei trademarks are the property 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 Cloud 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.

Huawei Cloud Computing Technologies Co., Ltd.

Address: Huawei Cloud Data Center Jiaoxinggong Road
Qianzhong Avenue
Gui'an New District
Gui Zhou 550029
People's Republic of China

Website: https://www.huaweicloud.com/intl/en-us/

Contents

1 Inspection Guide

1.1 Scenario

The purpose of routine inspection is to reduce potential system risks, to ensure long-term and stable running of the system, to reduce maintenance cost, and to ensure that the system properly processes services.

1.2 Inspection Items

MRS inspection items include:

- Alarm check: Check alarms on the ManageOne alarm platform and handle these alarms.
- Inspection on the management and control plane: Use FusionCare on ManageOne Operation Portal to inspect the MRS management and control plane.
- Inspection on the tenant plane: Use FusionInsight Tool MRS Prober to inspect the MRS tenant plane.

  Clusters on the MRS tenant plane are classified into ECS/BMS clusters (including ECS + passthrough disks, BMS, ECS/BMS storage-compute decoupling) and physical machine management clusters. Each type of clusters needs to be inspected after FusionInsight Tool MRS Prober is installed. For details, see **Tenant Plane Inspection**.

1.3 Inspection Frequency

The following table lists the recommended frequency for system inspection.

Table 1-1 Inspection frequency

Inspection Object	Inspection Item	Frequency
Management and control plane	Using FusionCare to perform automated acceptance of cloud services	After the installation, deployment and capacity expansion of cloud services
	Alarm check	One day
	Using FusionCare to inspect the management and control plane	One week
Tenant plane	Alarm check	One day
	Using FusionInsight Tool MRS Prober to inspect clusters on the tenant plane	One week

1.4 Management Plane Inspection

1.4.1 Routine Inspection Items

Table 1-2 lists the MRS inspection items supported by FusionCare on the management and control plane.

Table 1-2 Check items

No.	Item ID	Inspection Item	Inspection Method	Task Scenario	Standard
1	84000	Checking the CPU usage	FusionCare	• Routine health check • Pre-upgrade check	1. Log in to the node to be checked as user **opsadmin** and switch to user **root**. 2. Run the following command to check the CPU usage in the last three times: **top -n 3 -b |grep Cpu |grep -v grep | awk -F id '{print $1}' | awk -F ',' '{print $4}' | awk -F '%' '{print $1}'** 3. Subtract the obtained three values from 100 respectively to obtain the CPU usage in the last three times. If the values are all lower than the threshold, the check is passed. Otherwise, the check fails.

No.	Item ID	Inspection Item	Inspection Method	Task Scenario	Standard
2	84001	Checking the disk partition usage	FusionCare	• Routine health check • Pre-upgrade check	1. Log in to the node to be checked as user **opsadmin** and switch to user **root**. 2. Run the following command to check whether the usage of all partitions is less than the threshold: **df -h |grep -v "Use%" |awk '{print $6,$5}' | awk -F % '{print $1}'** If the usage of all partitions is lower than the threshold, the check is passed. Otherwise, the check fails.

No.	Item ID	Inspection Item	Inspection Method	Task Scenario	Standard
3	84002	Checking memory usage	FusionCare	• Routine health check • Pre-upgrade check	1. Log in to the node to be checked as user **opsadmin** and switch to user **root**. 2. Run the following command to check the memory usage in the last five seconds: **cat /proc/meminfo |grep "^MemTotal:|^MemFree:|^Buffers:|^Cached:|^SReclaimable:|^HugePages_Free:|^Hugepagesize:"** 3. Subtract other values from the value of **MemTotal** to obtain the occupied memory, and then calculate the memory usage. If the memory usage is lower than the threshold, the check is passed. Otherwise, the check fails.

No.	Item ID	Inspection Item	Inspection Method	Task Scenario	Standard
4	84003	Checking the large file list	FusionCare	• Routine health check • Pre-upgrade check	1. Log in to the node to be checked as user **opsadmin** and switch to user **root**. 2. Run the following commands: **find /var/log -type f -size +1G** **find /tmp -type f -size +1G** If no file larger than 1 GB is found, the check is passed. Otherwise, the check fails.

No.	Item ID	Inspection Item	Inspection Method	Task Scenario	Standard
5	84004	Checking whether the API container service is running properly	FusionCare	• Routine health check • Pre-upgrade check	1. Log in to the node to be checked as user **opsadmin** and switch to user **root**. 2. Run the following command to check whether the container service status is **Running**: **kubectl get pod -n mrs -owide | grep mrsapigw** Check whether all container services are in the **Running** state. If any container service is not in the **Running** state, the check fails.

No.	Item ID	Inspection Item	Inspection Method	Task Scenario	Standard
6	84005	Checking whether the Deploy service is running properly	FusionCare	• Routine health check • Pre-upgrade check	1. Log in to the node to be checked as user **opsadmin** and switch to user **root**. 2. Run the following command to check whether the container service status is **Running**: **kubectl get pods -n mrs -o wide | grep mrsdeployer** Check whether all container services are in the **Running** state. If any container service is not in the **Running** state, the check fails.

No.	Item ID	Inspection Item	Inspection Method	Task Scenario	Standard
7	84006	Checking whether the ZooKeeper service process is running properly	FusionCare	• Routine health check • Pre-upgrade check	1. Log in to the node to be checked as user **opsadmin** and switch to user **root**. 2. Run the following command to check whether the ZooKeeper service process exists: **lsof -i:2181** If the process does not exist, the check fails.

No.	Item ID	Inspection Item	Inspection Method	Task Scenario	Standard
8	84007	Checking whether the disk of the container node is read-only	FusionCare	• Routine health check • Pre-upgrade check	1. Log in to the node to be checked as user **opsadmin** and switch to user **root**. 2. Run the following command to check whether the read/write mode of the disk exists: **mount | grep "mnt"** If yes, the check fails.

No.	Item ID	Inspection Item	Inspection Method	Task Scenario	Standard
9	84008	Checking whether the system entropy is normal.	FusionCare	• Routine health check • Pre-upgrade check	1. Log in to the node to be checked as user **opsadmin** and switch to user **root**. 2. Run the following command to check the system entropy: **cat /proc/sys/kernel/random/entropy_avail** If the value is less than 1,000, the check fails.

No.	Item ID	Inspection Item	Inspection Method	Task Scenario	Standard
10	84099	Check The Integrity Of MRS Cloud Service Tree	FusionCare	• Routine health check • Pre-upgrade check	Check whether the MRS cloud service tree on ManageOne CMDB is complete and whether the VM agents on the service tree are normal.

1.4.2 Checking MRS Alarms

Check alarms on ManageOne Maintenance Portal, analyze the cause of each alarm in the alarm list and their impact on the system, and provide solutions based on the analysis result.

Prerequisites

You have logged in to ManageOne Maintenance Portal. For details, see **Logging In to ManageOne Maintenance Portal**.

Procedure

Step 1 Log in to the ManageOne Maintenance Portal and click **Alarm Management** > **Alarms** > **Current Alarms** in the navigation pane on the top.

Step 2 View the alarms.

1. Click **Filter** in the upper left corner and click **Location Info**. The **Location Info** dialog box is displayed.
2. Set **Operator** to **contains** and **Value** to **MRS**, and click **OK**. Click **OK** again to filter MRS alarms.

- If an alarm exits, go to **Step 3**.
- If no alarm exists, no further action is required.

Step 3 Locate the alarms.

In the search result list, click an alarm to view the details. In the alarm details page that is displayed, locate the fault based on the information listed in **Table 1-3**.

Table 1-3 Alarm information

Key Value	Description
Alarm serial number	Used to locate the fault.
IP address/URL/ Domain name	Used to locate the host generating the alarm.
Location	Provides information about the tenant ID, volume ID, and fault scope.
Additional Information	Provides remarks and the alarm handling history, helping find the troubleshooting personnel and track the handling status.

Step 4 Clear the alarms.

For details, see the alarm online help or "Alarm Handling" in *MapReduce Service (MRS) 3.3.1-LTS Fault Management (for Huawei Cloud Stack 8.3.1) in the MapReduce Service (MRS) 3.3.1-LTS Maintenance Guide (for Huawei Cloud Stack 8.3.1)*.

----End

1.4.3 FusionCare-based Inspections

1.4.3.1 Performing Routine Inspections

1.4.3.1.1 Creating a Health Check Task

This section describes how to create a routine inspection task for a project or site to identify potential risks to the system and to prevent failures. You also need to handle detected issues based on handling suggestions provided in the check results.

Prerequisites

- You have logged in to ManageOne Maintenance Portal. For details, see **Logging In to ManageOne Maintenance Portal**.
- The MRS node has been automatically added to the FusionCare environment information list.
  a. On the **Home** page of ManageOne Maintenance Portal, choose **FusionCare(Inspection)** > *Name of the region to be checked* in the **Common Links** area to go to the FusionCare system.
  b. Choose **System Management** > **Environment Configuration** from the main menu and check whether the MRS node information is displayed in the node list.
     The MRS node information is automatically added after the ManageOne information is added to FusionCare.

Procedure

Step 1 On ManageOne Maintenance Portal, click **O&M** > **Health Check**.

Step 2 Click **Create**. The **Create Task** page is displayed.

Step 3 Configure parameters based on **Table 1-4**.

Table 1-4 Task configuration information

Parameter	Description	Example
Task Name	Name of a health check task	Beijing_site

Parameter	Description	Example
Task Scenario	Scenario where the health check task is executed. The options are as follows: • **Routine health check**: Check basic check items required for routine O&M. • **Pre-upgrade check**: Before the upgrade, check whether the system status meets the upgrade requirements.	Routine health check
Task Policy	Select an execution policy of the current task. • **Real-Time Task**: The health check task is triggered immediately if no other health check task is in progress. • **Scheduled Task**: The health check task is executed at a specified period. If you select this option, set **Execution Time**. • **Periodic Task**: The health check task is periodically executed at a specified time on one or several days every week. If you select this option, set **Execution Time** to **By week** or **By month**.	Real-Time Task
Send check report via email	If you select this option, the task execution information is sent to administrators by email. If this option is enabled, configure the email by referring to **FusionCare Email Configuration Management**.	Send check report via email

Parameter	Description	Example
Customer Cloud	The desired customer cloud where the health check task is executed.	-
Select Objects	Target node where the health check task is executed. To execute a health check task, select at least one node. **NOTE** The object to be checked is the host housing the docker where the MRS management and control process is located, but not the MRS cluster provisioned by the tenant.	MRS
Select Check Items	Items of the health check task. To execute a health check task, select at least one health check item.	All items

Step 4 Click **Create Now**.

You can view the created health check task in the task list.

----End

1.4.3.1.2 Viewing a Health Check Task

Administrators can view health check tasks to learn about the status and execution progress of health check tasks in a timely manner.

Prerequisites

- You have logged in to ManageOne Maintenance Portal. For details, see **Logging In to ManageOne Maintenance Portal**.
- A health check task has been created. For details about how to create a health check task, see **Creating a Health Check Task**.

Procedure

Step 1 Choose **O&M** > **Health Check**. The health check task list is displayed.

Step 2 Check the items based on the check item names. **Table 1-5** describes the task details.

Table 1-5 Health check task details

Parameter	Description	Example
Name	Name of a health check task	beijing
Start Time	Time when the latest health check task is started	2021/03/01 21:23:06
Execution Duration	Time spent for executing the health check task	1m50s
Status	Health check task status. The options are as follows: ● **Finished**: A check task is complete. ● **Executing**: A check task is being executed. ● **Not started**: A check task is not started.	Finished
Progress	Execution progress of the current task	100%
Task Scenario	Health check task type. The options are as follows: ● Routine health check ● Pre-upgrade check	Routine health check
Task Policy	Execution policy of the current task. The options are as follows: ● Real-time task ● Scheduled task ● Periodic task	Real-Time Task
Object Check Pass Rate	Node check pass rate	100%
Check Item Pass Rate	Check item pass rate	100%
Created	User who creates the health check task	admin

----End

Related Operations

- Modify a health check task.

On the **Task List** page, locate the row that contains a desired task, and click **Modify** in the **Operation** column to modify the health check task. Currently, only scheduled and periodic tasks can be modified.

- Delete a health check task.

  On the **Task List** page, select a desired health check task, and above the task list, click **Delete**. On the displayed window, confirm the deletion of the health check task.

- Export a health check report.

  a. On the **Task List** page, locate the row that contains a desired task, and click **Export Report** in the **Operation** column.

  b. Choose a report type **Basic Report** or **Synthesis Report**.

  NOTE

  - A synthesis report contains the health check results of FusionCompute, FusionSphere OpenStack, ManageOne, and IaaS Service. The report file is in Word format.
  - **Customer Name** and **Signature** must be set for the synthesis report.

  c. Click **OK** to export the report.

1.4.3.1.3 Viewing Health Check Results

This section describes how to view faults and handling suggestions.

Prerequisites

- You have logged in to ManageOne Maintenance Portal.
- The health check task is complete.

Procedure

Step 1 Choose **O&M** > **Health Check**.

Step 2 Click the name of a finished health check task.

Step 3 On the task details page, perform the operations described in **Table 1-6**.

Table 1-6 Operations on the task details page

Task Name	Related Operations
Viewing basic information about a task	In the **Basic Information** area, view the name and status of the current task.
Viewing the object check pass rate and check item pass rate	The node check pass rate and check item pass rate are displayed in pie charts. You can select **By environment** or **By product**.

Task Name	Related Operations
Viewing component check results/tenant check results	• Viewing the details of a faulty task 1. On the **Component Check Result** or **Tenant Check Result** tab page, click the **Check Item Fault Details** tab and view the node where the fault is located displayed in the **Object Name** column. 2. Click the faulty node in the **Object Name** column. The health check result for this node is displayed. 3. Click a link in the **Check Item ID** column to view the troubleshooting suggestions for the check item. • Checking the status of each node in the task 1. Click **Details** in the **Operation** column to switch to the **Object Details** dialog box and view the results of the check items selected for the node in the task. 2. Click a link in the **Check Item ID** column to view the troubleshooting suggestions for the check item. **NOTE** The node status is determined based on the check results of the checked items. The node status can be normal or abnormal. If all check items pass the health check, the node status is normal. If a node contains at least one failed check item, the node status is abnormal. • Rechecking a task 1. Click **Retry** in the **Operation** column of the node. In the displayed dialog box, click **OK** to recheck the task. 2. Click **Details** in the **Operation** column of the node. In the displayed dialog box of the object details, click **Retry** in the **Operation** column to recheck the item.
Export a health check report.	1. Click **Export Report** in the upper right corner of the page. 2. Choose a report type **Basic Report** or **Synthesis Report**. **NOTE** A synthesis report contains the health check results of FusionCompute, FusionSphere OpenStack, ManageOne, and IaaS Service. The report file is in Word format. 3. Click **OK** to export the report.

----End

Related Operations

- Rechecking a task:

  In the upper right corner of the task details page, click **Retry** to recheck the task.
- Deleting a task: In the upper right corner of the task details page, click **Delete** to delete the task.

1.4.3.2 Performing Automated Acceptance

This section describes how to create an automated acceptance task for a project or site. Automated acceptance enables automatic commissioning, generates commissioning reports, and displays results to customers, reducing the manual acceptance test workload.

Logging In to FusionCare

Step 1 Log in to ManageOne Maintenance Portal using a browser as a system administrator.

- URL: **https://***Domain name of ManageOne Maintenance Portal*:**31943**. Alternatively, access **https://***Address for accessing the ManageOne unified portal* to log in to the ManageOne main portal and choose OperationCenter to access ManageOne Maintenance Portal.

  Obtain the domain name of ManageOne Maintenance Portal on the "2.3 Portal" sheet of the deployment parameter table exported from HCC Turnkey during the installation of Huawei Cloud Stack basic services.
- Login using a password: Enter the username and password.

  Default account: bss_admin. To obtain the default password, contact the system administrator or refer to the default password of the ManageOne Maintenance Portal account on the "Type A (Portal)" sheet in ***Huawei Cloud Stack 8.3.1 Account List*** .

  NOTE

  For ManageOne upgraded from 8.2.0 or earlier, the preset username is **admin**.
- Login using a USB key: Insert a USB key with preset user certificates, select the required device and certificate, and enter a PIN.

Step 2 On the **Home** page of ManageOne Maintenance Portal, choose **FusionCare(Inspection)** > *Name of the region to be checked* in the **Common Links** area to go to the FusionCare system.

Step 3 Choose **Automated Acceptance** from the main menu of FusionCare.

Step 4 In the navigation pane on the left, choose **Overview** to view the task progress, task result, recent task execution information, and historical acceptance statistics.

----End

Creating an Automated Acceptance Task

Step 1 Choose **Automated Acceptance** from the main menu of FusionCare.

Step 2 In the navigation pane on the left, choose **Acceptance Tasks**.

Step 3 On the **Acceptance Tasks** page, click **Create** in the upper left corner. The **Create Acceptance Task** page is displayed.

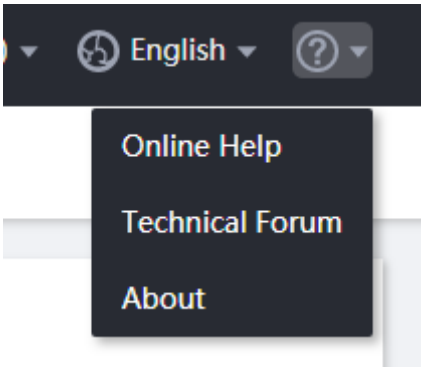
Step 4 In the **Basic Information** area, enter the following information:

- **Task Name**: Enter 1 to 16 characters. Only letters, digits, and underscores (_) are allowed. The name cannot be null.
- **Task Scenario**: Select the task scenario, which can be **Deployment**, **Upgrade**, or **Capacity expansion**.

  NOTE

  Currently, only **Deployment** and **Upgrade** are available for **Task Scenario**.
- **Customer Cloud**: Select the corresponding customer cloud.
- (Optional) **Enable snapshots**: To enable this function, you need to set **Server IP** and **Server Port**.

  Use Google Chrome to enable the screenshot function. For details about how to configure related parameters, choose **Online Help** in the upper right corner.



Step 5 In the **Objects and Cases** area, configure the following parameters:

- **Select Objects**: Select **MRS**.
- **Select Cases**: By default, all cases are selected.

Step 6 Configure parameters.

- Configuring parameters on the GUI

  Public parameters

  a. Click the **All cloud services** drop-down list.

  b. In the **Parameters** area, enter parameter values in the **Value** column based on the actual cluster information.

  c. Select **All cloud services** and click **Verify**.

  d. In the **Verification Result** column of the **Parameters** area, check whether all parameters pass the verification. Private parameters can be verified only after public parameters pass the verification.

  Private parameters

  a. Click the **MRS** drop-down list.

  b. In the **Parameters** area, enter parameter values in the **Value** column based on the actual cluster information.

c. Select **MRS** and click **Verify**.

NOTE

- Private parameters can be verified only after public parameters are verified.
- Post-upgrade acceptance does not involve private parameters, but you must click **Verify**.

d. In the **Verification Result** column of the **Parameters** area, check whether all parameters pass the verification.

- Configuring parameters using a template

a. Click **Export Template** to download the template to your local host. Open the template, enter parameter values in the **Value** column on the ManageOne and MRS sheets based on the actual cluster information, and save the template.

b. Click **Import Parameter**, select the template saved in **Step 6.a**, and click **Upload**.

c. Select **All cloud services** and click **Verify**. In the **Verification Result** column of the **Parameters** area, check whether all public parameters pass the verification. Private parameters can be verified only after public parameters pass the verification.

d. Select **MRS** and click **Verify**. In the **Verification Result** column of the **Parameters** area, check whether all private parameters pass the verification.

NOTE

If any parameter fails to pass the verification, check whether the parameter value is correct. If the value is incorrect, enter a correct value and verify the parameter again until it passes the verification.

Step 7 Click **Create Now**.

----End

Viewing the Execution Status of an Automated Acceptance Task

Step 1 Choose **Automated Acceptance** from the main menu of FusionCare.

Step 2 In the navigation pane on the left, choose **Acceptance Tasks**.

Step 3 On the **Acceptance Tasks** page, you can view the created acceptance task and the task status.

Parameter	Description
Task Name	Indicates the name of a created task.
Status	Includes **Executing**, **Finished**, or **Not started**.
Start Time	Indicates the task start time.
Execution Duration	Indicates the task execution time.

Parameter	Description
Progress	Indicates the task execution progress.
Task Scenario	Indicates in which scenario the acceptance task has been executed. Value options are **Deployment** and **Upgrade**.
Failed/Total Cases	Indicates the number of failed cases and total number of executed cases.
Created	Indicates the user who creates the task.
Operation	**Retry**: re-executes the acceptance task. **Export Report**: exports the case execution report.

----End

Viewing Acceptance Task Details

Step 1 Choose **Automated Acceptance** from the main menu of FusionCare.

Step 2 In the navigation pane on the left, choose **Acceptance Tasks**.

Step 3 On the **Acceptance Tasks** page, click an acceptance task name. The acceptance task details page is displayed.

Basic Information: includes the task name, status, scenario, start time, end time, and duration. You can export the acceptance report, and re-execute or delete the acceptance task.

Acceptance Result:

- **Cloud Service**: You can view the acceptance test status of cloud services.
- Case Execution Result: You can view the status of cloud acceptance test cases.
- Acceptance task results:

  Click **Case Details** in the **Operation** column. On the case details page that is displayed, you can view details about each acceptance case, including the acceptance objective, prerequisites, procedure, results, remarks, and snapshots.

  Click the drop-down button on the left of a failed task to view the case ID, name, and acceptance result.

Acceptance Details:

includes the status of **Pre-execution Cases**, **To-Be-Executed Cases**, and **Post-execution Cases**.

Parameter	Description
ID	You can click a case ID to go to the online help page of the case.
Name	Indicates the name of an acceptance case.
Cloud Service	Indicates the cloud service to which an acceptance case belongs.
Acceptance Result	Indicates the result of an acceptance case.
Operation	You can click **Case Details** in the **Operation** column to view the process information and case details.

----End

Exporting and Viewing a Task Report

Step 1 Choose **Automated Acceptance** from the main menu of FusionCare.

Step 2 In the navigation pane on the left, choose **Acceptance Tasks**.

Step 3 Click **Export Report** in the **Operation** column of the task whose acceptance report needs to be exported.

Step 4 After the download is complete, decompress the package on the local host, and open the report to view the execution results of each check case.

----End

1.4.4 Common Issues

1.4.4.1 Node Cannot Be Connected Due to the Special Characters in the Password When Interconnecting FusionCare with MRS

Symptom

When a user enters the password of user **root** on the **Add Node** page for interconnecting FusionCare with MRS, the special character **$** is contained in the password. As a result, an exception occurs and the node cannot be connected.

Solution

Add backslashes (\) before the special characters (including **$**, **"**, ****, **\n**, and **\r**) in the password for conversion before entering.

For example, convert **$** in the password **aaa$bbb** of user **root** to **\$**, and then enter the password **aaa\$bbb**.

1.4.4.2 FusionCare Email Configuration Management

This section guides technical support engineers and maintenance engineers to configure the mailbox. If the email forwarding function for the health check result is enabled, the FusionCare system will send the result to the specified email address.

Prerequisites

- You have logged in to ManageOne Maintenance Portal.
- The email server has been deployed by the user.

Procedure

Step 1 On the **Home** page of ManageOne Maintenance Portal, click **FusionCare(Inspection)** > *ID of the region to be checked* in the **Common Links** area to go to the FusionCare system.

Step 2 Choose **System Management** > **Email Configuration**. The **Email Configuration** page is displayed.

Step 3 Configure the email as shown in **Table 1-7**.

Figure 1-1 Email configuration

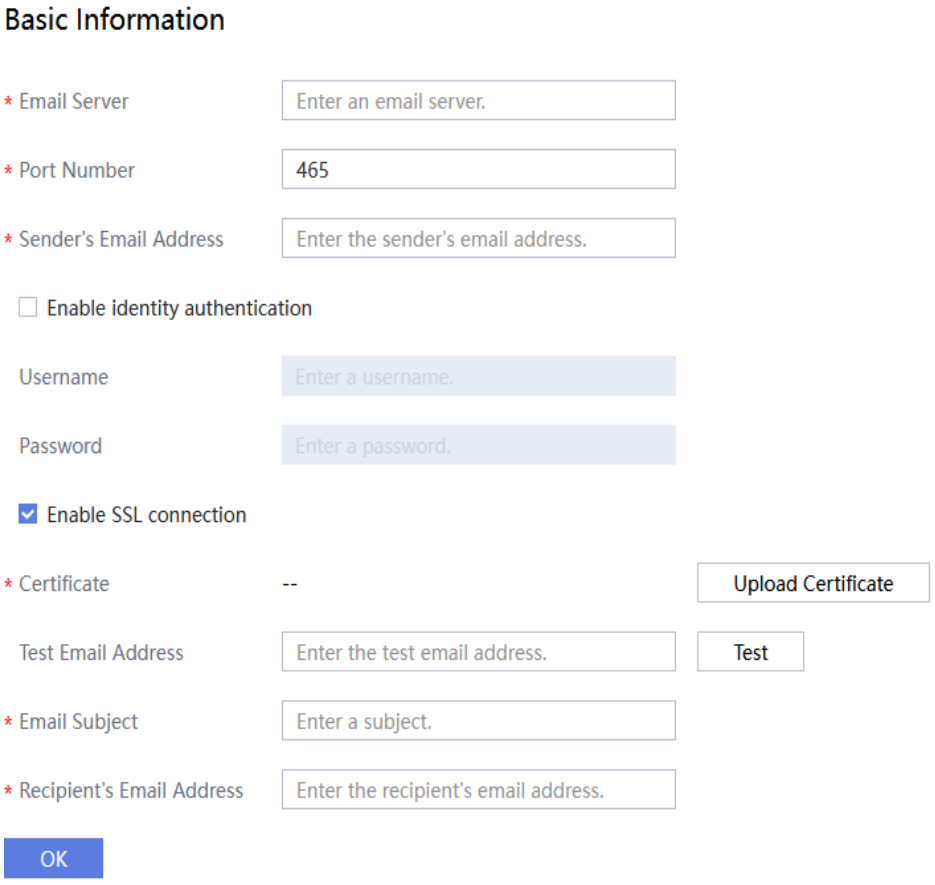


Table 1-7 Email configuration parameters

Parameter	Description
Email Server	Email server. Example: **192.168.190.73** or **mail.huawei.com**
Port Number	Port number of the email server
Sender's Email Address	Email address of the sender
Enable identity authentication	If this parameter is selected, you need to enter the username and password.
UserName	This parameter is mandatory when **Enable identity authentication** is selected.
Password	This parameter is mandatory when **Enable identity authentication** is selected.
Enable SSL connection	If the email server supports the SSL protocol, select **Enable SSL Connection** to improve system security.
Certificate	This parameter is displayed when **Enable SSL connection** is selected. Upload a security certificate in CRT format of the third-party email server.
Test Email Address	Set **Test Email Address** and click **Test**.
Email Subject	Email subject
Recipient's Email Address	Email address for receiving emails. Separate multiple email addresses with commas (,).

Step 4 Click **OK** to complete the email configuration.

NOTE

- Only the information about one email server can be added.
- If you need to modify the email information, perform the preceding steps again.

----End

1.4.4.3 MRS Is Not Displayed in the Inspection Object List During Inspection Task Creation

Symptom

When an inspection task is created, no MRS node can be selected in the inspection object list.

Solution

Step 1 Log in to FusionCare and choose **System Management > Environment Configuration**.

Step 2 Check whether the ManageOne and MRS nodes exist in the cloud node list of the current customer.

You do not need to manually add information about nodes on the MRS management plane. The system automatically adds information about VMs on the MRS management plane when interconnecting with ManageOne.

If the current node list does not contain the ManageOne node, click **Add Node** next to the current customer cloud name to manually add the ManageOne node.

Set related parameters as prompted. After the ManageOne node is added, wait for a period and check that the MRS node is displayed.

----End

1.5 Tenant Plane Inspection

1.5.1 Inspecting ECS Cluster/BMS Clusters

1. Download the tenant cluster inspection tool and *FusionInsight Tool MRS Prober XXX Inspection Guide* by referring to **Obtaining FusionInsight Tool MRS Prober**, and perform inspection on the BMS/ECS clusters by referring to the related documents.

   NOTE

   To inspect an MRS ECS/BMS cluster, you need to manually install FusionCare. If the local PC cannot directly access nodes in the cluster, you need to bind elastic IP addresses to the ECS/BMS nodes to upload the software package to the nodes in the cluster.

2. Check the cloud services on which MRS depends by referring to the inspection guides of other cloud services.

   In the storage-compute decoupling scenario, check the running status of OBS 3.0 by referring to sections related to the inspection guide.

   In the BMS cluster scenario, check the running status of the BMS service by referring to sections related to the inspection guide in *Bare Metal Server (BMS) 8.3.1 Maintenance Guide (for Huawei Cloud Stack 8.3.1)*.

1.5.2 Inspecting Physical Machine Clusters

Download the inspection tool and *FusionInsight Tool MRS Prober XXX Inspection Guide* by referring to **Obtaining FusionInsight Tool MRS Prober**, and perform inspection on a physical machine cluster by referring to the related documents.

1.5.3 Obtaining FusionInsight Tool MRS Prober

The FusionInsight Tool MRS Prober corresponding to the cluster version is required for routine inspection of MRS clusters on the tenant side. To obtain the tool, perform the following steps:

- Enterprise users:

  Use a Support account to log in to **https://support.huawei.com/enterprise/en/cloud-computing/fusioninsight-tool-pid-21624171/**

software, click the **Software Download** tab, and find and download FusionInsight Tool MRS Prober of the latest version and the corresponding inspection guide in the **Public Patch in V and R Version** area based on the release time.

- Carrier users:

  Use a Support account to log in to **https://support.huawei.com/carrier/docTypeNewOffering?col=product&path=PBI1-21430725/PBI1-21430757/PBI1-21431665/PBI1-21624171&resTab=SW**, click the **Software** tab, and find and download FusionInsight Tool MRS Prober of the latest version and the corresponding inspection guide in the **Version or Patch No.** area based on the release time.

Obtain the following software and documents:

- **SysChecker**_*XXX*.**zip**: SysChecker inspection service installation package
- **FusionCare** *XXX*.**zip**: FusionCare installation package
- *FusionInsight Tool MRS Prober XXX Inspection Guide*: tool usage guide

2 Backup and Restoration

2.1 Overview

2.1.1 Introduction

Overview

MRS supports manual backup or unified backup of MRS database data using ManageOne.

FusionInsight Manager supports the backup of system data and user data by components. The system can back up Manager data, component metadata, and service data.

Data can be backed up to local disks (LocalDir), local HDFS (LocalHDFS), remote HDFS (RemoteHDFS), NAS (NFS/CIFS), Object Storage Service (OBS) (supported only in ECS/BMS clusters), and SFTP server (SFTP). For details, see **Backing Up Metadata and Other Data**.

For a component that supports multiple services, multiple instances of a service can be backed up and restored. The backup and restoration operations are consistent with those of a service instance.

NOTE

- If you need to back up the cluster data to a third-party server, prepare the server and ensure that the server can communicate with the MRS cluster. For a physical machine cluster, the third-party server must be able to communicate with the cluster service plane. For an ECS/BMS cluster, the third-party server must be able to communicate with the VPC where the cluster is located. You can use Direct Connect to establish data connections between your local servers and VPCs on the cloud. For details, see *Direct Connect 8.3.1 Usage Guide (for Huawei Cloud Stack 8.3.1)*.
- The current restoration task cannot verify the service topology changes. You need to manually check whether the backup package used for restoration meets the requirements. If the backup data before scale-out or scale-in is used to restore data, the data will be restored to the status before scale-out or scale-in.

Backup and restoration tasks are performed in the following scenarios:

- Routine backup is performed to ensure the data security of the system and components.

- If the system is faulty, the data backup can be used to recover the system.
- If the active cluster is completely faulty, an image cluster identical to the active cluster needs to be created. You can use the backup data to restore the active cluster.

Principles

Task

Before backup or restoration, you need to create a backup or restoration task and set task parameters, such as the task name, backup data source, and type of the directory for storing backup files. Then you can execute the tasks to back up or restore data. When Manager is used to restore the data of HDFS, HBase, Elasticsearch, Hive, and NameNode, the cluster cannot be accessed.

Each backup task can back up data of different data sources and generate an independent backup file for each data source. All the backup files generated in a backup task form a backup file set, which can be used in restoration tasks. Backup data can be stored on Linux local disks, local cluster HDFS, and standby cluster HDFS.

Backup tasks support full backup and incremental backup policies. Cloud data backup tasks do not support incremental backup. If the backup directory type is NFS or CIFS, incremental backup is not recommended. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

NOTE

Task execution rules:

- If a task is being executed, the task cannot be executed repeatedly and other tasks cannot be started at the same time.
- The interval at which a periodic task is automatically executed must be greater than 120s. Otherwise, the task is postponed and will be executed in the next period. Manual tasks can be executed at any interval.
- When a periodic task is to be automatically executed, the current time cannot be 120s later than the task start time. Otherwise, the task is postponed and executed in the next period.
- When a periodic task is locked, it cannot be automatically executed and needs to be manually unlocked.
- Before an OMS, DBService, Kafka, MOTService, or NameNode backup task starts, ensure that the LocalBackup partition on the active management node has not less than 20 GB of available space. Otherwise, the backup task cannot be started.

When planning backup and restoration tasks, select the data to be backed up or restored strictly based on the service logic, data store structure, and database or table association. By default, the system creates periodic backup tasks **default-oms** and **default-*cluster ID*** at an interval of one hour. OMS metadata and cluster metadata, such as DBService and NameNode, can be fully backed up to local disks.

Snapshot

The system uses the snapshot technology to quickly back up data. Snapshots include HBase, Elasticsearch, and HDFS snapshots.

- HBase snapshots

  An HBase snapshot is a backup file of HBase tables at a specified time point. This backup file does not replicate service data or affect the RegionServer. The HBase snapshot replicates table metadata, including table descriptor, region info, and HFile reference information. The metadata can be used to restore data before the snapshot creation time.

- HDFS snapshots

  An HDFS snapshot is a read-only backup of HDFS at a specified time point. The snapshot is used in data backup, misoperation protection, and disaster recovery scenarios.

  The snapshot function can be enabled for any HDFS directory to create the related snapshot file. Before creating a snapshot for a directory, the system automatically enables the snapshot function for the directory. Creating a snapshot does not affect any HDFS operation. A maximum of 65,536 snapshots can be created for each HDFS directory.

  When a snapshot is being created for an HDFS directory, the directory cannot be deleted or modified before the snapshot is created. Snapshots cannot be created for the upper-layer directories or subdirectories of the directory.

- Elasticsearch snapshots

  An Elasticsearch snapshot uses the index data policy (snapshot API) of the backup cluster provided by Elasticsearch. The status and data of the current cluster at a specified time are backed up and saved to a specified snapshot repository. Your first snapshot will be a complete copy of data, and all subsequent snapshots will save the differences between the existing snapshots and the new data.

DistCp

Distributed copy (DistCp) is a tool used to replicate a large amount of data in HDFS in a cluster or between the HDFSs of different clusters. In a backup or restoration task of HBase, HDFS, Elasticsearch, or Hive, if you back up the data to HDFS of the standby cluster, the system invokes DistCp to perform the operation. Install the MRS software of the same version for the active and standby clusters and install the cluster.

DistCp uses MapReduce to implement data distribution, troubleshooting, restoration, and report. DistCp specifies different Map jobs for various source files and directories in the specified list. Each Map job copies the data in the partition that corresponds to the specified file in the list.

If you use DistCp to replicate data between HDFSs of two clusters, configure the cross-cluster mutual trust (mutual trust does not need to be configured for clusters managed by the same FusionInsight Manager) and cross-cluster replication for both clusters. When backing up the cluster data to HDFS in another cluster, you need to install the Yarn component. Otherwise, the backup fails.

Local rapid restoration

After using DistCp to back up the HBase, HDFS, and Hive data of the local cluster to the HDFS of the standby cluster, the HDFS of the local cluster retains the backup data snapshots. You can create local rapid restoration tasks to restore data by using the snapshot files in the HDFS of the local cluster.

NAS

Network Attached Storage (NAS) is a dedicated data storage server which includes the storage components and embedded system software. It provides the cross-platform file sharing function. By using NFS (supporting NFSv3 and NFSv4) and CIFS (supporting SMBv2 and SMBv3), you can connect the service plane of MRS to the NAS server to back up data to the NAS or restore data from the NAS.

NOTE

- Before data is backed up to the NAS, the system automatically mounts the NAS shared address to a local partition of the backup task execution node. After the backup is complete, the system unmounts the NAS shared partition from the backup task execution node.
- To prevent backup and restoration failures, do not access the shared address where the NAS server has been mounted to, for example, **/srv/BigData/LocalBackup/nas**, during data backup and restoration.
- When service data is backed up to the NAS, DistCp is used.
- On EulerOS 2.1, data cannot be backed up to or restored from NAS if NFS is used.

Specifications

Table 2-1 Specifications of the backup and restoration feature

Item	Specification
Maximum number of backup or restoration tasks	100
Number of concurrent tasks in a cluster	1
Maximum number of waiting tasks	199
Maximum size (GB) of backup files on a Linux local disk	600

NOTE

If service data is stored in the ZooKeeper upper-layer components, ensure that the number of znodes in a single backup or restoration task is not too large. Otherwise, the task will fail, and the ZooKeeper service performance will be affected. To check the number of znodes in a single backup or restoration task, perform the following operations:

- Ensure that the number of znodes in a single backup or restoration task is smaller than the upper limit of OS file handles. Specifically:
  1. To check the upper limit at the system level, run the **cat /proc/sys/fs/file-max** command.
  2. To check the upper limit at the user level, run the **ulimit -n** command.
- If the number of znodes in the parent directory exceeds the upper limit, back up and restore data in its sub-directories in batches. To check the number of znodes using ZooKeeper client scripts, perform the following operations:
  1. On **Homepage** of FusionInsight Manager, choose **Cluster** > **Services** > **ZooKeeper**. Click **Instance** and view the management IP address of each ZooKeeper role instance.
  2. Log in to the node where the client is located and run the following command:

     zkCli.sh -server *ip*:*port*, where, *ip* can be any management IP address, and the default port number is 24002.
  3. If the following information is displayed, login to the ZooKeeper server is successful:
     ```
     WatchedEvent state:SyncConnected type:None path:null
     [zk: ip:port(CONNECIED) 0]
     ```
  4. Run the **getusage** command to check the number of znodes in the directory to be backed up.

     For example, **getusage /hbase/region**. In the command output, **Node count=xxxxxx** indicates the number of znodes stored in the **region** directory.

Table 2-2 Specifications of the default task

Item	OMS	Elasticsearch	HBase	IoTDB	ClickHouse	Kafka	DBService	Flink	NameNode
Backup period	1 hour								
Maximum number of backups	168 (7-day historical data)								24 (one-day historical data)
Maximum size of a backup file	10 MB	20 MB	10 MB	10 MB	20 MB	512 MB	100 MB	1 GB	20 GB
Maximum size of disk space used	1.64 GB	3.28 GB	1.64 GB	1.64 GB	3.28 GB	84 GB	16.41 GB	168 GB	480 GB
Storage path of backup data	*Data storage path*/**LocalBackup/** of the active and standby management nodes								

NOTE

- When periodic backup is performed for HDFS, Hive, Elasticsearch, and HBase, snapshots are created for protected directories. Affected by the snapshot mechanism, deleting data between two backups does not release disk space immediately.
- The backup data of the default backup task must be periodically transferred and saved outside the cluster based on the enterprise O&M requirements.
- Administrators can create DistCp backup tasks to save OMS, DBService, and NameNode data to external clusters.
- The execution time of a cluster data backup task can be calculated using the following formula: Task execution time = Volume of data to be backed up/Network bandwidth between the cluster and the backup device. In practice, you are advised to multiply the calculated time by 1.5 to get the reference value of the task execution time.
- Executing a data backup task affects the maximum I/O performance of the cluster. Therefore, you are advised to execute a backup task during off-peak hours.

2.1.2 Backup and Restoration Policies

MRS supports the backup of Manager configuration data, component metadata or other data, and service data of specific components based on service requirements.

Table 2-3 Manager configuration data to be backed up

Backup Type	Backup Content	Backup Directory Type
OMS	Database data (excluding alarm data) and configuration data in the cluster management system by default	• LocalDir • LocalHDFS • RemoteHDFS • NFS • CIFS • SFTP • OBS

Table 2-4 Component metadata or other data to be backed up

Backup Type	Backup Content	Backup Directory Type
DBService	Metadata of the components (including Loader, Hive, Spark, Oozie, CDL, Redis, and Hue) managed by DBService. For a cluster with multiple services installed, back up the metadata of multiple Hive and Spark service instances.	• LocalDir • LocalHDFS • RemoteHDFS • NFS • CIFS • SFTP • OBS

Backup Type	Backup Content	Backup Directory Type
Flink	Flink metadata.	• LocalDir • LocalHDFS • RemoteHDFS
Kafka	Kafka metadata.	• LocalDir • LocalHDFS • RemoteHDFS • NFS • CIFS • OBS
NameNode	HDFS metadata. After multiple NameServices are added, backup and restoration are supported for all of them and the operations are consistent with those of the default hacluster instance.	• LocalDir • RemoteHDFS • NFS • CIFS • SFTP • OBS
Yarn	Information about the Yarn service resource pool.	
HBase	**tableinfo** files and data files of HBase system tables.	
Solr	Solr metadata.	• LocalDir • LocalHDFS • NFS • CIFS • SFTP
Redis	Redis service data.	• LocalHDFS
IoTDB	IoTDB metadata.	• LocalDir • NFS • RemoteHDFS • CIFS • SFTP
ClickHouse	ClickHouse metadata.	• LocalDir • RemoteHDFS • OBS
Containers	Containers metadata.	• LocalDir • LocalHDFS • RemoteHDFS

Backup Type	Backup Content	Backup Directory Type
RTDService	RTDService metadata.	• LocalDir • LocalHDFS • RemoteHDFS

Table 2-5 Service data of specific components to be backed up

Backup Type	Backup Content	Backup Directory Type
HBase	Table-level user data. For a cluster with multiple services installed, backup and restoration are supported for multiple HBase service instances and the backup and restoration operations are consistent with those of a single HBase service instance.	• RemoteHDFS • NFS • CIFS • SFTP • OBS
HDFS	Directories or files of user services. **NOTE** Encrypted directories cannot be backed up or restored.	• RemoteHDFS • NFS • CIFS • SFTP
Hive	Table-level user data. For a cluster with multiple services installed, backup and restoration are supported for multiple Hive service instances and the backup and restoration operations are consistent with those of a single Hive service instance.	
Elasticsearch	Index data. For a cluster with multiple services installed, backup and restoration are supported for multiple Elasticsearch service instances and the backup and restoration operations are consistent with those of a single Elasticsearch service instance.	• RemoteHDFS • NFS
Solr	Index data. For a cluster with multiple services installed, backup and restoration are supported for multiple Solr service instances and the backup and restoration operations are consistent with those of a single Solr service instance.	• RemoteHDFS
IoTDB	IoTDB service data.	• RemoteHDFS
ClickHouse	Table-level user data.	• RemoteHDFS • OBS

Backup Type	Backup Content	Backup Directory Type
MOTService	MOTService service data.	• RemoteHDFS

Note that some components do not provide data backup or restoration:

- Kafka supports replicas and allows multiple replicas to be specified when a topic is created.
- CDL data is stored in DBService and Kafka. A system administrator can create DBService and Kafka backup tasks to back up data.
- MapReduce and Yarn data is stored in HDFS. Therefore, they rely on the backup and restoration provided by HDFS.
- Backup and restoration of service data in ZooKeeper are performed by their own upper-layer components.

2.2 Backup

2.2.1 Backing Up Data on the Management and Control Plane

Manually Backing Up the MRS Database Data

The administrator can manually log in to an MRS-DB node to back up the MySQL database. If manual backup is required, perform the following steps to back up the **datasightemr** and **mrs** databases of MRS:

Step 1 Log in to ManageOne Maintenance Portal as the system administrator. In the **Common Links** navigation tree, click Service_OM and select a region to go to the Service OM page.

Step 2 Choose **Services** > **Resource** > **Compute Resource** from the main menu.

Step 3 Click the **VMs** tab, enter a keyword in the search box to search for the VM name, for example, **MRS_DB**, and record the IP address of the VM.

Step 4 Use PuTTY to log in to the active MRS-DB node as user **opsadmin** and run the following command to switch to user **root**:

su - root

NOTE

See **Determining the Active/Standby Status of MRS-DB Nodes** for details about how to check the active/standby status of the MRS-DB node and the floating IP address of the database.

For details about the default password, see ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator.

Step 5 Run the following commands to back up the database to a specified directory that has the operation permission:

su - mysql

/data/mysql/base/bin/mysqldump -udatasightemr -h*MRS-DB floating IP address* **-P7306 -p***Password of database user datasightemr* **--flush-logs --master-data=2 --single-transaction --set-gtid-purged=OFF --databases datasightemr >** */tmp/mrs_datasightemr_backup.sql*

/data/mysql/base/bin/mysqldump -udatasightemr -h*MRS-DB floating IP address* **-P7306 -p***Password of database user datasightemr* **--flush-logs --master-data=2 --single-transaction --set-gtid-purged=OFF --databases mrs >** */tmp/mrs_mrs_backup.sql*

Contact O&M personnel to obtain the default password of database user **datasightemr**.

----End

Backing Up the MRS Database Data on ManageOne in a Unified Manner

The MRS-DB database can be connected to the SFTP server through ManageOne for unified backup. For details, see "Maintenance Guide" > "O&M Guide" > "Backup and Restoration" > "Data Backup" > "Backup Management" in *Huawei Cloud Stack 8.3.1 Product Documentation* . When you select the backup object, choose the **S-MRS-MySQL** option under the CloudDB system category for the MRS-DB database.

NOTE

If MRS 3.0.2-LTS.2 or an earlier version is upgraded to the current version and MRS MySQL is not registered with CloudCMDB, perform the operations in **Solution to the Situation Where MRS-MySQL Is Not Registered with CloudCMDB**.

2.2.2 Backing Up Data on the Tenant Plane

2.2.2.1 Backing Up OMS Data

Scenario

To ensure data security of FusionInsight Manager, you need to back up FusionInsight Manager data, especially before and after a critical operation (such as capacity expansion and reduction) on FusionInsight Manager. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up Manager data. Both automatic and manual backup tasks are supported.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust

has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.

- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path*/**LocalBackup/** has sufficient space on the active and standby management nodes.
- If you want to back up data to NAS, you have deployed the NAS server in advance.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Set **Backup Object** to **OMS**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-6 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.

Parameter	Description
Backup Policy	• **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times** **NOTE** • Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. • If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, select **OMS**.

Step 7 Set **Path Type** of **OMS** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files.

  The default storage directory is *Data storage path***/LocalBackup/**, for example, **/srv/BigData/LocalBackup**.

  If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select this option, set the following parameters:

  – **Target Path**: indicates the HDFS directory for storing the backup files. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.

  – **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

  – **Cluster for Backup**: Enter the cluster name mapping to the backup directory.

  – **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  – **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in

remote cluster of the cluster, or the NameService name of a configured remote cluster.

- **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
- **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
- **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
- **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **Source Cluster**: Select the cluster of the Yarn queue used by the backup data.
- **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the source cluster.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)

  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Server Shared Path**: indicates the backup path on the SFTP server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **OBS**: indicates that backup files are stored in OBS.

  If you select this option, set the following parameters:
  - **Target Path**: indicates the OBS directory for storing backup data.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files.

The format of the backup file name is *Version_Data source_Task execution time*.**tar.gz**.

----End

2.2.2.2 Backing Up Metadata and Other Data

2.2.2.2.1 Backing Up CDL Data

Scenario

To ensure CDL service data security, you need to back up CDL data especially before a major operation on CDL (such as upgrade or migration). The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

CDL data is stored in DBService and Kafka. You can create DBService and Kafka backup tasks on FusionInsight Manager to back up CDL data. Both automatic and manual backup tasks are supported.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path*/**LocalBackup/** has sufficient space on the active and standby management nodes.
- If you want to back up data to NAS, you have deployed the NAS server in advance.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-7 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.

Parameter	Description
Backup Policy	• **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times** **NOTE** • Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. • If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 Set **Configuration** to **DBService** and **Kafka**.

NOTE

If there are multiple DBService or Kafka services, all DBService or Kafka services are backed up by default. You can click **Assign Service** to specify the services to be backed up.

Step 7 Set **Path Type** of **DBService** to a backup directory type. For details about how to set the parameters, see **Step 7**.

Step 8 Set **Path Type** of **Kafka** to a backup directory type. For details about how to set the parameters, see **Step 7**.

Step 9 Click **OK**.

Step 10 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files.

The format of the backup file name is *Version_Data source_Task execution time*.**tar.gz**.

----End

2.2.2.2.2 Backing Up ClickHouse Metadata

Scenario

To ensure ClickHouse metadata security or before a major operation (such as upgrade or migration), you need to back up ClickHouse metadata. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up ClickHouse metadata. Both automatic and manual backup tasks are supported.

NOTICE

ClickHouse metadata cannot be backed up in the federation scenario.

Before backup, ensure that all ClickHouseServer nodes are running properly. Otherwise, the backup may fail.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path*/**LocalBackup/** has sufficient space on the active and standby management nodes.
- If the active and standby clusters are deployed in security mode and they are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active and standby clusters are deployed in normal mode, no mutual trust is required.
- The time on the active and standby clusters must be the same, and the NTP service on the active and standby clusters uses the same time source.
- In the active/standby cluster, if data is remotely backed up to HDFS, ensure that the value of **HADOOP_RPC_PROTECTION** of ClickHouse is the same as that of **hadoop.rpc.protection** of HDFS.
- When data is remotely backed up to HDFS, HDFS encrypted directories are not supported.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task. **Periodic** indicates that the backup task is periodically executed. **Manual** indicates that the backup task is manually executed.

To create a periodic backup task, set the following parameters:

- **Started**: indicates the time when the task is started for the first time.
- **Period**: indicates the task execution interval. The options include **Hours** and **Days**.

- **Backup Policy**: Only **Full backup every time** is supported.

Step 6 In **Configuration**, select **ClickHouse** under **Metadata and other data**.

Step 7 Set **Path Type** of **ClickHouse** to a backup directory type.

Table 2-8 Path of backup data

Directory Type	Description
LocalDir	Indicates that the backup files are stored on the local disk of the active management node, and the standby management node automatically synchronizes the backup files. The default storage directory is *Data storage path*/**LocalBackup/**, for example, **/srv/BigData/LocalBackup**. If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.
RemoteHDFS	Indicates that the backup files are stored in the HDFS directory of the standby cluster. Only the latest backup file can be retained. Historical backup files are overwritten. You also need to configure the following parameters: • **Destination NameService Name**: indicates the NameService name of the standby cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster. • **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**. • **Destination Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster. • **Destination Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster. • **Destination NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the destination cluster. • **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.

Directory Type	Description
OBS	Indicates that backup files are stored in an OBS directory. • **Target Path**: indicates the OBS directory for storing backup data. • **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files. The format of the backup file name is *Data source_Task execution time***.tar.gz**.

----End

2.2.2.2.3 Backing Up Containers Metadata

Scenario

To ensure Containers metadata security or before a major operation on Containers (such as upgrade or migration), you need to back up Containers metadata. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up Containers metadata. Both automatic and manual backup tasks are supported.

NOTICE

Containers metadata cannot be backed up in the federation scenario.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path***/LocalBackup/** has sufficient space on the active and standby management nodes.

- If the active cluster is deployed in security mode (with Kerberos authentication enabled) and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode (with Kerberos authentication disabled), mutual trust is not required.
- The time on the active and standby clusters must be the same, and the NTP service on the active and standby clusters uses the same time source.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task. **Periodic** indicates that the backup task is periodically executed. **Manual** indicates that the backup task is manually executed.

To create a periodic backup task, set the following parameters:

Table 2-9 Periodic backup parameters

Parameter	Description
Started	The time when the task is started for the first time.
Period	The task execution interval. Value options include **Hours** and **Days**.
Backup Policy	• **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times**

Step 6 In **Configuration**, select **Containers** under **Metadata and other data**.

Step 7 Set **Path Type** of **Containers** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files.

  The default storage directory is *Data storage path*/**LocalBackup/**, for example, **/srv/BigData/LocalBackup**.

  If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select this option, you also need to configure the following parameters:

  - **Target Path**: indicates the HDFS directory for storing the backup files. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **RemoteHDFS**: indicates that backup files are stored in HDFS of the standby cluster.

  You also need to set the following parameters:

  - **Destination NameService Name**: indicates the NameService name of the standby cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Destination Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
  - **Destination Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
  - **Destination NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the destination cluster.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files. The format of the backup file name is *Data source_Task execution time*.**tar.gz**.

----End

2.2.2.2.4 Backing Up DBService Data

Scenario

To ensure DBService service data security, you need to back up DBService data, especially before a major operation on DBService (such as upgrade or migration).

The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up DBService data. Both automatic and manual backup tasks are supported.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path*/**LocalBackup/** has sufficient space on the active and standby management nodes.
- If you want to back up data to NAS, you have deployed the NAS server in advance.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-10 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.

Parameter	Description
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	• **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times** **NOTE** • Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. • If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, select **DBService**.

NOTE

If there are multiple DBService services, all DBService services are backed up by default. You can click **Assign Service** to specify the services to be backed up.

Step 7 Set **Path Type** of **DBService** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files.

  The default storage directory is *Data storage path*/**LocalBackup/**, for example, **/srv/BigData/LocalBackup**.

  If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.
- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select this option, set the following parameters:

  – **Target Path**: indicates the HDFS directory for storing the backup files. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.

  – **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

  – **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  - **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the source cluster.
- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)

  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Server Shared Path**: indicates the backup path on the SFTP server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **OBS**: indicates that backup files are stored in OBS.

  If you select this option, set the following parameters:
  - **Target Path**: indicates the OBS directory for storing backup data.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files.

The format of the backup file name is *Version_Data source_Task execution time*.**tar.gz**.

----End

2.2.2.2.5 Backing Up Flink Metadata

Scenario

To ensure Flink metadata security or before a major operation on Flink (such as upgrade or migration), you need to back up Flink metadata. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up Flink metadata. Both automatic and manual backup tasks are supported.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- The HDFS and Yarn services have been installed if data needs to be backed up to HDFS.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path*/**LocalBackup/** has sufficient space on the active and standby management nodes.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task. **Periodic** indicates that the backup task is periodically executed. **Manual** indicates that the backup task is manually executed.

To create a periodic backup task, set the following parameters:

- **Started**: indicates the time when the task is started for the first time.
- **Period**: indicates the task execution interval. The options include **Hours** and **Days**.
- **Backup Policy**: Only **Full backup every time** is supported.

Step 6 In **Configuration**, select **Flink** under **Metadata and other data**.

Step 7 Set **Path Type** of **Flink** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files.

  The default storage directory is *Data storage path*/**LocalBackup/**, for example, **/srv/BigData/LocalBackup**.

If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select this option, set the following parameters:

  - **Target Path**: indicates the HDFS directory for storing the backup files. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  - **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the source cluster.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files. The format of the backup file name is *Data source_Task execution time***.tar.gz**.

----End

2.2.2.2.6 Backing Up HBase Metadata

Scenario

To ensure HBase metadata security (including tableinfo files and HFiles) or before a major operation on HBase system tables (such as upgrade or migration), you need to back up HBase metadata to prevent HBase service unavailability caused by HBase system table directory or file damages. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up HBase metadata. Both automatic and manual backup tasks are supported.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path*/**LocalBackup/** has sufficient space on the active and standby management nodes.
- If you want to back up data to NAS, you have deployed the NAS server in advance.
- The **fs.defaultFS** parameter settings of HBase are the same as those of Yarn and HDFS.
- If HBase data is stored in the local HDFS, HBase metadata can be backed up to OBS. If HBase data is stored in OBS, data backup is not supported.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-11 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	• **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times** **NOTE** • Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. • If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, select **HBase** under **Metadata and other data**.

NOTE

If there are multiple HBase services, all HBase services are backed up by default. You can click **Assign Service** to specify the services to be backed up.

Step 7 Set **Path Type** of **HBase** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files.

  The default storage directory is *Data storage path*/**LocalBackup/**, for example, **/srv/BigData/LocalBackup**.

  If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.
- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  – **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in

remote cluster of the cluster, or the NameService name of a configured remote cluster.

- **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
- **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
- **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
- **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the source cluster.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Server Shared Path**: indicates the backup path on the SFTP server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **OBS**: indicates that backup files are stored in OBS.

  If you select this option, set the following parameters:

  - **Target Path**: indicates the OBS directory for storing backup data.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files. The format of the backup file name is *Version_Data source_Task execution time*.**tar.gz**.

----End

2.2.2.2.7 Backing Up IoTDB Metadata

Scenario

To ensure IoTDB metadata security and prevent the IoTDB service from being unavailable due to IoTDB metadata file damages, you need to back up IoTDB metadata. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up IoTDB metadata. Both automatic and manual backup tasks are supported.

Prerequisites

- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path*/**LocalBackup/** has sufficient space on the active and standby management nodes.
- If you want to back up data to NAS, you have deployed the NAS server in advance.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- Backup policies, including the backup task type, period, backup object, backup directory, and Yarn queue required by the backup task are planned based on service requirements.
- The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-12 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	Indicates a periodic backup policy. • **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times** **NOTE** Incremental backup is not supported when component metadata is backed up. Only **Full backup every time** is supported.

Step 6 In **Configuration**, select **IoTDB** under **Metadata and other data**.

NOTE

If there are multiple IoTDB services, all IoTDB services are backed up by default. You can click **Assign Service** to specify the services to be backed up.

Step 7 Set **Path Type** of **IoTDB** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files. By default, the backup files are stored in *Data storage path*/**LocalBackup/**.

  If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  - **Destination NameService Name**: indicates the NameService name of the standby cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Destination Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
  - **Destination Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
  - **Destination NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the standby cluster.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Server Shared Path**: indicates the backup path on the SFTP server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files.

The format of the backup file name is *Version_Data source_Task execution time*.**tar.gz**.

----End

2.2.2.2.8 Backing Up Kafka Metadata

Scenario

To ensure Kafka metadata security or before a major operation on ZooKeeper (such as upgrade or migration), you need to back up Kafka metadata. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up Kafka metadata. Both automatic and manual backup tasks are supported.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path*/**LocalBackup/** has sufficient space on the active and standby management nodes.
- If you want to back up data to NAS, you have deployed the NAS server in advance.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-13 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	• **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times** **NOTE** • Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. • If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, select **Kafka**.

NOTE

If there are multiple Kafka services, all Kafka services are backed up by default. You can click **Assign Service** to specify the services to be backed up.

Step 7 Set **Path Type** of **Kafka** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files. By default, the backup files are stored in *Data storage path*/**LocalBackup/**.

  If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.
- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select this option, set the following parameters:

  – **Target Path**: indicates the HDFS directory for storing the backup files. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.

  – **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

  – **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

If you select this option, set the following parameters:

- **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.
- **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
- **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
- **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
- **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory,

and the user group and owner group of the shared path must be **nobody:nobody**.)

- **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

- **OBS**: indicates that backup files are stored in OBS.

  If you select this option, set the following parameters:

  - **Target Path**: indicates the OBS directory for storing backup data.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files. The format of the backup file name is *Version_Data source_Task execution time*.**tar.gz**.

----End

2.2.2.2.9 Backing Up NameNode Data

Scenario

To ensure NameNode service data security, you need to back up NameNode data especially before a major operation on NameNode (such as upgrade or migration). The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up NameNode data. Both automatic and manual backup tasks are supported.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data*

storage path/**LocalBackup/** has sufficient space on the active and standby management nodes.

- If you want to back up data to NAS, you have deployed the NAS server in advance.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-14 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	Only **Full backup every time** is supported. **NOTE** • Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. • If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, select **NameNode**.

Step 7 Set **Path Type** of **NameNode** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files. By default, the backup files are stored in *Data storage path*/**LocalBackup/**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster. If you select this option, set the following parameters:
  - **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the service plane IP address of the NameNode in the standby cluster.
  - **Target Path**: indicates the path for storing backup files.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol. If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol. If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.

  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Server Shared Path**: indicates the backup path on the SFTP server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **OBS**: indicates that backup files are stored in OBS.

  If you select this option, set the following parameters:
  - **Target Path**: indicates the OBS directory for storing backup data.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files.

The format of the backup file name is *Version_Data source_Task execution time*.**tar.gz**.

----End

2.2.2.2.10 Backing Up Redis Data

Scenario

To ensure Redis data security, you need to back up Redis cluster data especially before a major operation on a Redis cluster (such as upgrade or migration). The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up Redis cluster data. To prevent the Redis service from being severely affected, manually back up data.

Prerequisites

- You have checked that HDFS services have been deployed in the current cluster.
- The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task. **Periodic** indicates that the backup task is periodically executed. **Manual** indicates that the backup task is manually executed.

To create a periodic backup task, set the following parameters:

- **Started**: indicates the time when the task is started for the first time.
- **Period**: indicates the task execution interval. The options include **Hours** and **Days**.
- **Backup Policy**: Only **Full backup every time** is supported.

  NOTE

  - AOF persistency is performed on full data when a backup task is executed on Redis. If the service data volume is large, the performance is greatly affected. You are advised not to periodically back up Redis data.
  - Manually back up Redis data during off-peak hours.

Step 6 In **Configuration**, select **Redis**.

Step 7 Set **Path Type** of **Redis** to a backup directory type.

The following backup directory types are supported:

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select this option, set the following parameters:

  – **Target Path**: indicates the HDFS directory for storing the backup files. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.

  NOTE

  - The target path of Redis data backup cannot be set to the SM4 encrypted partition of HDFS.
  - The target path can be a path that does not exist in HDFS. A path will be automatically created during the backup.

  – **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

  – **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for the backup task in the backup directory. The format of the subdirectory name is *Backup task name_Data source_Task creation time*.

----End

2.2.2.2.11 Backing Up RTDService Metadata

Scenario

To ensure RTDService metadata security or before a major operation on RTDService (such as upgrade or migration), you need to back up RTDService metadata. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up RTDService metadata. Both automatic and manual backup tasks are supported.

NOTICE

RTDService metadata cannot be backed up in the federation scenario.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path*/**LocalBackup/** has sufficient space on the active and standby management nodes.
- If the active cluster is deployed in security mode (with Kerberos authentication enabled) and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode (with Kerberos authentication disabled), mutual trust is not required.
- The time on the active and standby clusters must be the same, and the NTP service on the active and standby clusters uses the same time source.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task. **Periodic** indicates that the backup task is periodically executed. **Manual** indicates that the backup task is manually executed.

To create a periodic backup task, set the following parameters:

Table 2-15 Periodic backup parameters

Parameter	Description
Started	The time when the task is started for the first time.
Period	The task execution interval. Value options include **Hours** and **Days**.
Backup Policy	• **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times**

Step 6 In **Configuration**, select **RTDService** under **Metadata and other data**.

Step 7 Set **Path Type** of **RTDService** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files.

  The default storage directory is *Data storage path*/**LocalBackup/**, for example, **/srv/BigData/LocalBackup**.

  If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select this option, you also need to configure the following parameters:

  - **Target Path**: indicates the HDFS directory for storing the backup files. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **RemoteHDFS**: indicates that backup files are stored in HDFS of the standby cluster.

  You also need to set the following parameters:

  - **Destination NameService Name**: indicates the NameService name of the standby cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Destination Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
  - **Destination Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
  - **Destination NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the destination cluster.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data

source backup files. The format of the backup file name is *Data source_Task execution time***.tar.gz**.

----End

2.2.2.2.12 Backing Up Solr Metadata

Scenario

Solr metadata is stored in ZooKeeper. To ensure Solr metadata security or before a major operation on ZooKeeper (such as upgrade or migration), you need to back up Solr metadata. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up Solr metadata. Both automatic and manual backup tasks are supported.

Prerequisites

- The backup type, period, policy, and other specifications have been planned based on the service requirements and you have checked whether *Data storage path***/LocalBackup/** has sufficient space on the active and standby management nodes.
- If you want to back up data to NAS, you have deployed the NAS server in advance.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-16 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.

Parameter	Description
Backup Policy	Indicates a periodic backup policy. • **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times** **NOTE** Incremental backup is not supported when component metadata is backed up. Only **Full backup every time** is supported.

Step 6 In **Configuration**, select **Solr** under **Metadata and other data**.

NOTE

If there are multiple Solr services, all Solr services are backed up by default. You can click **Assign Service** to specify the services to be backed up.

Step 7 Set **Path Type** of **Solr** to a backup directory type.

The following backup directory types are supported:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node and the standby management node automatically synchronizes the backup files. By default, the backup files are stored in *Data storage path*/**LocalBackup/**.

  If you select this option, you need to set the maximum number of replicas to specify the number of backup file sets that can be retained in the backup directory.
- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:
  - **Destination NameService Name**: indicates the NameService name of the standby cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.

  - **Destination Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
  - **Destination Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
  - **Destination NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the standby cluster.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Server Shared Path**: indicates the backup path on the SFTP server.

- **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Task creation time*, and the subdirectory is used to save data source backup files.

The format of the backup file name is *Version_Data source_Task execution time*.**tar.gz**.

----End

2.2.2.3 Backing Up Service Data

2.2.2.3.1 Backing Up ClickHouse Service Data

Scenario

To ensure ClickHouse service data security, you need to back up ClickHouse service data, especially before a major operation on ClickHouse (such as upgrade or migration). The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up ClickHouse service data. Both automatic and manual backup tasks are supported.

NOTICE

ClickHouse service data cannot be backed up in the federation scenario.

Before backup, ensure that all ClickHouseServer nodes are running properly. Otherwise, the backup may fail.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active and standby clusters are deployed in security mode and they are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active and standby clusters are deployed in normal mode, no mutual trust is required.
- The time on the active and standby clusters must be the same, and the NTP service on the active and standby clusters uses the same time source.

- You have planned the backup type, period, object, and directory based on service requirements.
- The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.
- In the active/standby cluster, if data is remotely backed up to HDFS, ensure that the value of **HADOOP_RPC_PROTECTION** of ClickHouse is the same as that of **hadoop.rpc.protection** of HDFS.
- When data is remotely backed up to HDFS, HDFS encrypted directories are not supported.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-17 Periodic backup parameters

Parameter	Description
Started	The time when the task is started for the first time.
Period	The task execution interval. Value options include **Hours** and **Days**.
Backup Policy	• **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times** **NOTE** • Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported.

Step 6 In **Configuration**, choose **Service Data** > **ClickHouse** > **ClickHouse**.

Step 7 Set **Path Type** of **ClickHouse** to a backup directory type.

Table 2-18 Path of backup data

Directory Type	Description
RemoteHDFS	Indicates that the backup files are stored in the HDFS directory of the standby cluster. Only the latest backup file can be retained. Historical backup files are overwritten. You also need to configure the following parameters: • **Destination NameService Name**: indicates the NameService name of the standby cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster. • **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**. • **Destination Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster. • **Destination Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster. • **Destination NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the destination cluster. • **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
OBS	Indicates that backup files are stored in an OBS directory. • **Target Path**: indicates the OBS directory for storing backup data. • **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

Step 8 Set **Maximum Number of Recovery Points** to any value from **1** to **1000** because this parameter is not used by ClickHouse.

Step 9 Set **Backup Content** to one or multiple ClickHouse tables to be backed up.

You can select backup data using either of the following methods:

- Adding a backup data file
  a. Click **Add**.

b. Select the table to be backed up under **File Directory**, and click **Add** to add the table to **Backup Content**.

c. Click **OK**.

- Selecting using regular expressions

a. Click **Query Regular Expression**.

b. Enter the logical cluster and database to which the ClickHouse table belongs in the first text box as prompted. The logical cluster and database must match the existing logical cluster and database, for example, **/default_cluster/database**.

c. Enter a regular expression in the second box. Standard regular expressions are supported. For example, to search for all tables that contain the keyword **test** in the database, enter **test.***.

d. Click **Refresh** to view the displayed tables in **Directory Name**.

e. Click **Synchronize** to save the result.

NOTE

- When entering regular expressions, click + or − to add or delete an expression.
- If the selected table or directory is incorrect, click **Clear Selected Node** to deselect it.

Step 10 Click **Verify** to check whether the backup task is configured correctly.

The possible causes of the verification failure are as follows:

- The target NameNode IP address is incorrect.
- The directory or table to be backed up does not exist.
- The name of the NameService is incorrect.

Step 11 Click **OK**.

Step 12 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Data source_Task creation time*, and the subdirectory is used to save latest data source backup files.

----End

2.2.2.3.2 Backing Up Doris Data

Scenario

To ensure Doris service data security, you need to back up Doris service data especially before you perform a major operation on Doris (such as upgrade or migration). The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create automatic or manual tasks on FusionInsight Manager to back up Doris data.

Prerequisites

- To back up data to a remote HDFS, the following conditions must be met:
  - A standby cluster for backing up data has been created. The authentication mode must be the same as that of the active cluster.
  - At least one DBroker instance of the Doris service has been deployed in the active cluster.
  - If the active and standby clusters are deployed in security mode and they are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active and standby clusters are deployed in normal mode, no mutual trust is required.
  - The time on the active and standby clusters must be the same, and the NTP service on the active and standby clusters uses the same time source.
  - The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.
  - The value of **hadoop.rpc.protection** of Doris must be the same as that of **hadoop.rpc.protection** of HDFS in both active and standby clusters.
- You have planned the backup type, period, object, and directory based on service requirements.
- If you want to back up data to OBS, you have connected the Doris cluster to OBS and have the permission to access OBS. For details, see "Configuring Interconnection Between Doris and OBS" in MapReduce Service (MRS) 3.3.1-LTS User Guide (for Huawei Cloud Stack 8.3.1) in the MapReduce Service (MRS) 3.3.1-LTS Usage Guide (for Huawei Cloud Stack 8.3.1).

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

- **Periodic**: indicates that the backup task is periodically executed. If you select this mode, you need to set other parameters by referring to **Table 2-19**.
- **Manual**: indicates that the backup task is manually executed.

Table 2-19 Periodic backup parameters

Parameter	Description
Started	Time when the task is started for the first time
Period	The task execution interval. The options include **Hours** and **Days**.

Parameter	Description
Backup Policy	• **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times** **NOTE** Currently, Doris supports only **full backup each time**. Incremental backup is not supported.

Step 6 In **Configuration**, select **Doris** under **Metadata and other data**.

Step 7 Set **Path Type** of **Doris** to a backup directory type.

Table 2-20 Path of backup data

Directory Type	Description
RemoteHDFS	Indicates that backup files are stored in the HDFS directory of the standby cluster. If you select this option, configure the following parameters: • **Destination NameService Name**: indicates the NameService name of the standby cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS on the standby cluster. • **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**. • **Destination NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster. • **Destination NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS configuration of the destination cluster. • **DBroker IP**: indicates the IP address of a service plane where the DBroker role in the cluster is deployed. The DBroker is used to transmit data during backup. • **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
OBS	Indicates that backup files are stored in an OBS directory. You need to set the **Target Path** to the OBS directory for storing backup data.

Step 8 Set **Maximum Number of Recovery Points** to the number of snapshots that can be retained in the cluster.

Step 9 Set **Backup Content** to one or multiple Doris tables to be backed up.

You can select backup data using either of the following methods:

- Adding a backup data file
  a. Click **Add**.
  b. Select the table to be backed up under **File Directory**, and click **Add** to add the table to **Backup Content**.
  c. Click **OK**.
- Using regular expressions
  a. Click **Query Regular Expression**.
  b. Enter the database where the Doris tables are located in the first text box as prompted. The database must be the same as the existing database, for example, **/example_db**.
  c. Enter a regular expression in the second box. Standard regular expressions are supported. For example, to search for all tables that contain the keyword **test** in the database, enter **test.***.
  d. Click **Refresh** to view the displayed tables in **Directory Name**.
  e. Click **Synchronize** to save the result.

  NOTE

  - When entering regular expressions, click + or — to add or delete an expression.
  - If the selected table or directory is incorrect, click **Clear Selected Node** to deselect it.

Step 10 Click **Verify** to check whether the backup task is configured correctly.

The possible causes of the verification failure are as follows:

- The target NameNode IP address is incorrect.
- The name of the NameService is incorrect.
- The table to be backed up does not exist.
- The format of the table to be backed up is incorrect.
- The tables to be backed up must come from the same database.

Step 11 Click **OK**.

Step 12 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Data source_Task creation time*, and the subdirectory is used to save latest data source backup files. All the backup file sets are stored in the related snapshot directories.

----End

2.2.2.3.3 Backing Up Elasticsearch Service Data

Scenario

To ensure Elasticsearch service data security, you need to back up Elasticsearch service data especially before a major operation on Elasticsearch (such as upgrade or migration). The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up Elasticsearch service data. Both automatic and manual backup tasks are supported.

NOTICE

- During the snapshot creation, the search and query functions are not affected. After the snapshot creation process starts, new data is not recorded in the snapshot. Only one snapshot can be created at a time.
- When a backup task is created, only the indexes that have been enabled in the cluster are displayed as backup objects. The disabled indexes are not displayed on the GUI. This way, the disabled indexes are not backed up.
- If some indexes selected for a backup task are disabled before the backup task starts, the disabled indexes will not be backed up. Only enabled indexes are backed up. If all indexes are disabled, the backup task fails to be executed.
- The Elasticsearch service data backup needs to invoke the snapshot interface through the EsNode1 instance. Therefore, ensure that all EsNode1 instances in the cluster are in good health status and can receive requests normally. To ensure successful backup, do not perform operations such as adding, deleting, stopping, or restarting Elasticsearch instances, stopping or restarting the Elasticsearch service, or stopping or restarting the cluster.
- If a large amount of data needs to be backed up in the cluster, back up data at the index level in batches. Otherwise, the backup takes a long time.
- To prevent a large amount of data from being fully backed up each time, create a periodic backup task when creating an index. In this case, data is fully backed up in the first backup task, and incremental backup is performed in subsequent periodic backup tasks.
- If a backup task fails, log in to the backup directory of the target (**RemoteHDFS** and **NFS**), which is the value of **Target Path** for a backup to remote HDFS or the value of **Server Shared Path** for a backup to the NFS. Delete the subdirectory (*Backup task name_Data source_Task creation time*) corresponding to the backup task name to delete data that fails to be backed up.
- Before the backup, check whether the index to be backed up is in the green state and no shard is lost. Otherwise, the backup fails.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby

cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.

- The HDFS and Yarn services have been installed if data needs to be backed up to HDFS. For the Elasticsearch cluster in normal mode, service data cannot be backed up to HDFS in a cluster in security mode.
- The HDFS service has been installed if data needs to be backed up to the NAS.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The HDFS and NAS client in the standby cluster have sufficient space. You are advised to save backup files in a custom directory.
- When backing up the Elasticsearch service data to the NAS (NFS), you have deployed the NAS server and performed the following operations:

  After the NAS is started and a shared path is created, create a local repository path and mount it to the shared path of the NAS. For details about how to run commands in batches, see **How Do I Run Commands or Access Files on Multiple Nodes in a Cluster?**.

  a. Create a shared path of the NAS and change its owner and permission. For example, the shared path is **/var/nfs**.

     - Run **mkdir /var/nfs** to create a path.
     - Run **chown 65534:65534 /var/nfs** to change the owner.
     - Run **chmod 777 /var/nfs** to change the permission.

  b. On each server, run the following command to mount the local repository path to the shared path of the NAS:

     mount *ip*:**/var/nfs /***Data storage path*/**elasticsearch/nas**

     In the command, **ip** indicates the IP address of the NAS server. For example:

     mount *ip*:**/var/nfs /srv/BigData/elasticsearch/nas**

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task. **Periodic** indicates that the backup task is periodically executed. **Manual** indicates that the backup task is manually executed.

To create a periodic backup task, set the following parameters:

- **Started**: indicates the time when the task is started for the first time.
- **Period**: indicates the task execution interval. The options include **Hours** and **Days**.
- **Backup Policy**: indicates the volume of data to be backed up in each task execution. Only **Full backup at the first time and incremental backup subsequently** is supported.

Step 6 In **Configuration**, choose **Elasticsearch** > **Elasticsearch** under **Service data**.

Step 7 Set **Path Type** of **Elasticsearch** to a backup directory type. Elasticsearch data cannot be backed up to a directory encrypted using RangerKMS.

The following backup directory types are supported:

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Destination Hadoop PRC Mode**: indicates the value of **hadoop.rpc.protection** in the HDFS basic configuration of the destination cluster.
  - **Destination Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the destination cluster.
  - **Destination Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the destination cluster.
  - **Destination NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the destination cluster.
  - **Target Path**: indicates the HDFS directory for storing destination cluster backup data. The path cannot be an HDFS hidden directory, such as snapshot or recycle bin directory, or a default system directory.
- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Backup Speed of a Single Instance (MB/s)**: indicates the speed of backing up data for a single instance. The default value is **50 MB/s**. Change the backup speed based on the actual volume of backup data.
  - **Restoration Speed of a Single Instance (MB/s)**: indicates the speed of restoring data for a single instance. The default value is **50 MB/s**. Change the restoration speed based on the actual volume of backup data.
  - **Server Shared Path**: indicates the shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)

Step 8 Set **Maximum Number of Recovery Points** to any value from **1** to **1000** because this parameter is not used by Elasticsearch.

Step 9 Set **Backup Content** to one or more indexes to be backed up.

You can select backup data using either of the following methods:

- Adding a backup data file
  a. Click **Add**.
  b. Select the table to be backed up under **File Directory**, and click **Add** to add the table to **Backup Content**.
  c. Click **OK**.
- Selecting using regular expressions
  a. Click **Query Regular Expression**.
  b. Enter a regular expression in the second text box. Standard regular expressions are supported. For example, to get indexes containing **es**, enter **.*es.***. To get indexes starting with **es**, enter **es.***. To get indexes ending with **es**, enter **.*es**.
  c. Click **Refresh** to view the displayed tables in **Directory Name**.
  d. Click **Synchronize** to save the result.

  NOTE

  - When entering regular expressions, click + or − to add or delete an expression.
  - If the selected table or directory is incorrect, click **Clear Selected Node** to deselect it.

Step 10 Click **Verify** to check whether the backup task is configured correctly.

The possible causes of the verification failure are as follows:

- The destination active or standby NameNode IP address or port number is incorrect.
- The name of the index to be backed up does not exist in the Elasticsearch cluster.

Step 11 Click **OK**.

Step 12 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Data source_Task creation time*, and the subdirectory is used to save latest data source backup files. All the backup file sets are stored in the related snapshot directories.

----End

2.2.2.3.4 Backing Up HBase Service Data

Scenario

To ensure HBase service data security or before a major operation on HBase (such as upgrade or migration), you need to back up HBase service data. The backup

data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up HBase service data. Both automatic and manual backup tasks are supported.

The following situations may occur during the HBase service data backup:

- When a user creates an HBase table, **KEEP_DELETED_CELLS** is set to **false** by default. When the user backs up this HBase table, deleted data will be backed up and junk data may exist after data restoration. This parameter can be set to **true** manually when an HBase table is created based on service requirements.
- When a user manually specifies the timestamp when writing data into an HBase table and the specified time is earlier than the last backup time of the HBase table, new data may not be backed up in incremental backup tasks.
- The HBase backup function cannot back up the access control lists (ACLs) for reading, writing, executing, creating, and managing HBase global or namespaces. After HBase data is restored, you need to reset the role permissions on FusionInsight Manager.
- If the backup data of the standby cluster is lost in an existing HBase backup task, the next incremental backup will fail, and you need to create an HBase backup task again. However, the next full backup task will be normal.
- MRS can back up HBase data in a cluster in security mode to a third-party server outside the cluster. For details, see **Backing Up OMS Data to a Third-Party Server Outside a Cluster**.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- Backup policies, including the backup task type, period, backup object, backup directory, and Yarn queue required by the backup task are planned based on service requirements.
- The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.
- On the HDFS client, you have executed the **hdfs lsSnapshottableDir** command as user **hdfs** to check the list of directories for which HDFS snapshots have been created in the current cluster and ensured that the HDFS parent directory or subdirectory where data files to be backed up are stored

does not have HDFS snapshots. Otherwise, the backup task cannot be created.

- If you want to back up data to NAS, you have deployed the NAS server in advance.
- The **fs.defaultFS** parameter settings of HBase are the same as those of Yarn and HDFS.
- If HBase data is stored in the local HDFS, HBase service data can be backed up to OBS. If HBase data is stored in OBS, data backup is not supported.
- If you want to back up data to OBS, you have connected the current cluster to OBS and have the permission to access OBS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-21 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	• **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times** **NOTE** • Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. • If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, choose **HBase** > **HBase** under **Service data**.

Step 7 Set **Path Type** of **HBase** to a backup directory type.

The following backup directory types are supported:

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  - **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Server Shared Path**: indicates the backup path on the SFTP server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.

  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **OBS**: indicates that backup files are stored in OBS.

  If you select this option, you also need to configure the following parameters:
  - **Target Path**: indicates the OBS directory for storing backup data.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

Step 8 Set **Maximum Number of Recovery Points** to the number of snapshots that can be retained in the cluster.

Step 9 Set **Backup Content** to one or multiple HBase tables to be backed up.

You can select backup data using either of the following methods:

- Adding a backup data file
  a. Click **Add**.
  b. Select the table to be backed up under **File Directory**, and click **Add** to add the table to **Backup Content**.
  c. Click **OK**.
- Selecting using regular expressions
  a. Click **Query Regular Expression**.
  b. Enter the namespace where the HBase tables are located in the first text box as prompted. The namespace must be the same as the existing namespace, for example, **default**.
  c. Enter a regular expression in the second text box. Standard regular expressions are supported. For example, to get all tables in the namespace, enter **([\s\S]*?)**. To get tables whose names consist of letters and digits, for example, **tb1**, enter **tb\d***.
  d. Click **Refresh** to view the displayed tables in **Directory Name**.
  e. Click **Synchronize** to save the result.

  NOTE
  - When entering regular expressions, click + or − to add or delete an expression.
  - If the selected table or directory is incorrect, click **Clear Selected Node** to deselect it.

Step 10 Click **Verify** to check whether the backup task is configured correctly.

The possible causes of the verification failure are as follows:

- The target NameNode IP address is incorrect.

- The queue name is incorrect.
- The parent directory or subdirectory of the HDFS directory where HBase table data files to be backed up are stored has HDFS snapshots.
- The directory or table to be backed up does not exist.

Step 11 Click **OK**.

Step 12 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *xxx/Backup task name_Data source_Task creation time*, and the subdirectory is used to save latest data source backup files. All the backup file sets are stored in the related snapshot directories.

----End

2.2.2.3.5 Backing Up HDFS Service Data

Scenario

To ensure HDFS service data security, you need to back up HDFS service data, especially before a major operation on HDFS (such as upgrade or migration). The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up HDFS service data. Both automatic and manual backup tasks are supported.

NOTE

Encrypted directories cannot be backed up or restored.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- Backup policies, including the backup task type, period, backup object, backup directory, and Yarn queue required by the backup task are planned based on service requirements.

- The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.
- On the HDFS client, you have executed the **hdfs lsSnapshottableDir** command as user **hdfs** to check the list of directories for which HDFS snapshots have been created in the current cluster and ensured that the HDFS parent directory or subdirectory where data files to be backed up are stored does not have HDFS snapshots. Otherwise, the backup task cannot be created.
- If data needs to be backed up to NAS, you must ensure that the NAS server has been deployed and NAS disks can be mounted to each NodeManager node.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-22 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	• **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times** **NOTE** • Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. • If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, select **HDFS**.

Step 7 Set **Path Type** of **HDFS** to a backup directory type.

The following backup directory types are supported:

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, set the following parameters:

  - **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.

  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol. If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Server Shared Path**: indicates the backup path on the SFTP server.

- **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
- **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
- **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
- **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

Step 8 Set **Maximum Number of Recovery Points** to the number of snapshots that can be retained in the cluster.

Step 9 Set **Backup Content** to one or multiple HDFS directories to be backed up based on service requirements.

You can select backup data using either of the following methods:

- Adding a backup data file
  a. Click **Add**.
  b. Select the table to be backed up under **File Directory**, and click **Add** to add the table to **Backup Content**.
  c. Click **OK**.
- Selecting using regular expressions
  a. Click **Query Regular Expression**.
  b. Enter the parent directory full path of the directory in the first text box as prompted. The directory must be the same as the existing directory, for example, **/tmp**.
  c. Enter a regular expression in the second text box. Standard regular expressions are supported. For example, to get all files or subdirectories in the parent directory, enter **([\s\S]*?)**. To get files whose names consist of letters and digits, for example, **file*1***, enter **file\d***.
  d. Click **Refresh** to view the displayed directories in **Directory Name**.
  e. Click **Synchronize** to save the result.

  NOTE

  - When entering regular expressions, click + or − to add or delete an expression.
  - If the selected table or directory is incorrect, click **Clear Selected Node** to deselect it.
  - The backup directory cannot contain files that have been written for a long time. Otherwise, the backup task will fail. Therefore, you are not advised to perform operations on the top-level directory, such as **/user**, **/tmp**, and **/mr-history**.

Step 10 Click **Verify** to check whether the backup task is configured correctly.

The possible causes of the verification failure are as follows:

- The target NameNode IP address is incorrect.

- The queue name is incorrect.
- The parent directory or subdirectory of the HDFS directory where data files to be backed up are stored has HDFS snapshots.
- The directory or table to be backed up does not exist.
- The name of the NameService is incorrect.

Step 11 Click **OK**.

Step 12 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Data source_Task creation time*, and the subdirectory is used to save latest data source backup files. All the backup file sets are stored in the related snapshot directories.

----End

2.2.2.3.6 Backing Up Hive Service Data

Scenario

To ensure Hive service data security, you need to back up Hive service data especially before a major operation on Hive (such as upgrade or migration). The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up Hive service data. Both automatic and manual backup tasks are supported.

- Hive backup and restoration cannot identify the service and structure relationships of objects such as Hive tables, indexes, and views. When executing backup and restoration tasks, you need to manage unified restoration points based on service scenarios to ensure proper service running.
- Hive backup and restoration do not support Hive on RDB data tables. You need to back up and restore original data tables in external databases independently.
- If the backup data of the standby cluster is lost in an existing Hive backup task that contains Hive on HBase tables, the next incremental backup will fail, and you need to create a Hive backup task again. However, the next full backup task will be normal.
- After the backup function of FusionInsight Manager is used to back up the HDFS directories at the Hive table level, the Hive tables cannot be deleted and recreated.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.

- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- Backup policies, including the backup task type, period, backup object, backup directory, and Yarn queue required by the backup task are planned based on service requirements.
- The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.
- On the HDFS client, you have executed the **hdfs lsSnapshottableDir** command as user **hdfs** to check the list of directories for which HDFS snapshots have been created in the current cluster and ensured that the HDFS parent directory or subdirectory where data files to be backed up are stored does not have HDFS snapshots. Otherwise, the backup task cannot be created.
- If you want to back up data to NAS, you have deployed the NAS server in advance.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-23 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.

Parameter	Description
Backup Policy	• **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times** **NOTE** • Incremental backup is not supported when Manager data and component metadata are backed up. Only **Full backup every time** is supported. • If **Path Type** is set to **NFS** or **CIFS**, incremental backup cannot be used. When incremental backup is used for NFS or CIFS backup, the latest full backup data is updated each time the incremental backup is performed. Therefore, no new recovery point is generated.

Step 6 In **Configuration**, choose **Hive** > **Hive**.

Step 7 Set **Path Type** of **Hive** to a backup directory type.

The following backup directory types are supported:

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster. If you select this option, set the following parameters:
  - **Destination NameService Name**: indicates the NameService name of the standby cluster. You can set it to the NameService name (**haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**) of the built-in remote cluster of the cluster, or the NameService name of a configured remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Target NameNode IP Address**: indicates the IP address of the NameNode service plane in the standby cluster. It can be of an active or standby node.
  - **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol. If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol. If you select this option, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Server Shared Path**: indicates the configured shared directory of the NAS server. (The shared path of the server cannot be set to the root directory, and the user group and owner group of the shared path must be **nobody:nobody**.)
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Server Shared Path**: indicates the backup path on the SFTP server.
  - **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

Step 8 Set **Maximum Number of Recovery Points** to the number of snapshots that can be retained in the cluster.

Step 9 Set **Backup Content** to one or multiple Hive tables to be backed up.

You can select backup data using either of the following methods:

- Adding a backup data file
  a. Click **Add**.
  b. Select the table to be backed up under **File Directory**, and click **Add** to add the table to **Backup Content**.
  c. Click **OK**.
- Selecting using regular expressions
  a. Click **Query Regular Expression**.
  b. Enter the database where the Hive tables are located in the first text box as prompted. The database must be the same as the existing database, for example, **default**.
  c. Enter a regular expression in the second text box. Standard regular expressions are supported. For example, to get all tables in the database,

enter **([\s\S]*?)**. To get tables whose names consist of letters and digits, for example, **tb1**, enter **tb\d***.

d. Click **Refresh** to view the displayed tables in **Directory Name**.

e. Click **Synchronize** to save the result.

NOTE

- When entering regular expressions, click + or — to add or delete an expression.
- If the selected table or directory is incorrect, click **Clear Selected Node** to deselect it.

Step 10 Click **Verify** to check whether the backup task is configured correctly.

The possible causes of the verification failure are as follows:

- The target NameNode IP address is incorrect.
- The queue name is incorrect.
- The parent directory or subdirectory of the HDFS directory where data files to be backed up are stored has HDFS snapshots.
- The directory or table to be backed up does not exist.
- The name of the NameService is incorrect.

Step 11 Click **OK**.

Step 12 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Data source_Task creation time*, and the subdirectory is used to save latest data source backup files. All the backup file sets are stored in the related snapshot directories.

----End

2.2.2.3.7 Backing Up IoTDB Service Data

Scenario

To ensure IoTDB service data security, you need to back up IoTDB service data especially before a major operation on IoTDB (such as upgrade or migration). The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up IoTDB service data. Both automatic and manual backup tasks are supported.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster. Currently, IoTDB data can be backed up only to HDFS.

- For the IoTDB cluster in normal mode, service data cannot be backed up to HDFS in a cluster in security mode.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-24 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	Indicates a periodic backup policy. • **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times** **NOTE** Incremental backup is not supported when component service data is backed up. Only **Full backup every time** is supported.

Step 6 In **Configuration**, choose **IoTDB** > **IoTDB** under **Service data**.

Step 7 Set **Path Type** of **IoTDB** to a backup directory type.

The following backup directory types are supported:

RemoteHDFS: indicates that the backup files are stored in the HDFS directory of the standby cluster.

If you select this option, set the following parameters:

- **Destination NameService Name**: indicates the NameService name of the standby cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
- **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
- **Destination Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
- **Destination Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
- **Destination NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the standby cluster.
- **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.

Step 8 Set **Backup Content** to one or multiple service data records to be backed up.

You can select backup data using either of the following methods:

- Adding a backup data file
  a. Click **Add**.
  b. Select the table to be backed up under **File Directory**, and click **Add** to add the table to **Backup Content**.
  c. Click **OK**.
- Selecting using regular expressions
  a. Click **Query Regular Expression**.
  b. Enter the parent directory full path of the directory in the first text box as prompted. The directory must be the same as the existing directory, for example, **/root**.
  c. Enter a regular expression in the second text box. Standard regular expressions are supported. For example, to get all files or subdirectories in the parent directory, enter **([\s\S]*?)**. To get files whose names consist of letters and digits, for example, **file*1***, enter **file\d***.
  d. Enter a regular expression in the second text box. Standard regular expressions are supported. For example, to get objects containing **test**, enter **.*test.***. To get objects starting with **test**, enter **test.***. To get objects ending with **test**, enter **.*test**.
  e. Click **Refresh** to view the displayed directories in **Directory Name**.
  f. Click **Synchronize** to save the result.

NOTE

- When entering regular expressions, click + or — to add or delete an expression.
- If the selected table or directory is incorrect, click **Clear Selected Node** to deselect it.
- The backup directory cannot contain files that have been written for a long time. Otherwise, the backup task will fail. Therefore, you are not advised to perform operations on the top-level directory, such as **/user**, **/tmp**, and **/mr-history**.

Step 9 Click **Verify** to check whether the backup task is configured correctly.

The possible causes of the verification failure are as follows:

- The target NameNode IP address is incorrect.
- The data to be backed up does not exist.

Step 10 Click **OK**.

Step 11 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Data source_Task creation time*, and the subdirectory is used to save latest data source backup files. All the backup file sets are stored in the related snapshot directories.

----End

2.2.2.3.8 Backing Up MOTService Service Data

Scenario

To ensure MOTService service data security or before a major operation on MOTService (such as upgrade or migration), you need to back up MOTService service data. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up MOTService service data. Both automatic and manual backup tasks are supported.

NOTICE

MOTService service data cannot be backed up in the federation scenario.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster.
- If the active cluster is deployed in security mode (with Kerberos authentication enabled) and the active and standby clusters are not managed

by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode (with Kerberos authentication disabled), mutual trust is not required.

- The time on the active and standby clusters must be the same, and the NTP service on the active and standby clusters uses the same time source.
- You have planned the backup type, period, object, and directory based on service requirements.
- The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-25 Periodic backup parameters

Parameter	Description
Started	The time when the task is started for the first time.
Period	The task execution interval. Value options include **Hours** and **Days**.
Backup Policy	Only **Full backup every time** is supported. • **Full backup at the first time and incremental backup subsequently** • **Full backup every time** • **Full backup once every n times**

Step 6 In **Configuration**, choose **MOTService** > **MOTService** under **Service data**.

Step 7 Set **Path Type** of **MOTService** to a backup directory type.

Currently, the backup directory supports only the **RemoteHDFS** type. **RemoteHDFS** indicates the HDFS directory for storing backup files in the standby cluster.

- **Destination NameService Name**: indicates the NameService name of the standby cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.

- **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
- **Destination Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
- **Destination Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
- **Destination NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the destination cluster.
- **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.

Step 8 Click **OK**.

Step 9 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Data source_Task creation time*, and the subdirectory is used to save latest data source backup files.

----End

2.2.2.3.9 Backing Up Solr Service Data

Scenario

To ensure Solr service data security, you need to back up Solr service data especially before a major operation on Solr (such as upgrade or migration). The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

You can create a backup task on FusionInsight Manager to back up Solr service data. Both automatic and manual backup tasks are supported.

NOTICE

- During snapshot creation, the search and query functions are not affected. After the snapshot creation process starts, new data is not recorded in the snapshot. Only one snapshot can be created at a time.
- If some indexes selected during backup task creation are deleted before the backup task is started, the deleted indexes will not be backed up. If all indexes are deleted, the backup task fails to be executed.
- Ensure that the running status of all instances in the cluster is normal and can receive requests properly. To ensure successful backup, do not perform operations such as adding, deleting, stopping, or restarting Solr instances, stopping or restarting the Solr service, or stopping or restarting the cluster.
- If a large amount of data needs to be backed up in the cluster, back up data at the index level in batches. Otherwise, the backup takes a long time.
- If a backup task fails, log in to the backup directory of the target (**RemoteHDFS**), which is the value of **Target Path** for a backup to remote HDFS. Delete the subdirectory (*Backup task name_Data source_Task creation time*) corresponding to the backup task name to delete data that fails to be backed up.
- Before the backup, check whether the index to be backed up is in the green state and no shard is lost. Otherwise, the backup fails.

Prerequisites

- If data needs to be backed up to the remote HDFS, you have prepared a standby cluster for data backup. The authentication mode of the standby cluster is the same as that of the active cluster. For other backup modes, you do not need to prepare the standby cluster. Currently, Solr data can be backed up only to HDFS.
- For the Solr cluster in normal mode, service data cannot be backed up to HDFS in a cluster in security mode.
- If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The HDFS in the standby cluster has sufficient space. You are advised to save backup files in a custom directory.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 Click **Create**.

Step 3 Set **Name** to the name of the backup task.

Step 4 Select the desired cluster from **Backup Object**.

Step 5 Set **Mode** to the type of the backup task.

Periodic indicates that the backup task is executed by the system periodically. **Manual** indicates that the backup task is executed manually.

Table 2-26 Periodic backup parameters

Parameter	Description
Started	Indicates the time when the task is started for the first time.
Period	Indicates the task execution interval. The options include **Hours** and **Days**.
Backup Policy	Indicates the volume of data to be backed up in each task execution. Only **Full backup every time** is supported.

Step 6 In **Configuration**, choose **Solr** > **Solr** under **Service data**.

Step 7 Set **Path Type** of **Solr** to a backup directory type.

The following backup directory types are supported:

RemoteHDFS: indicates that the backup files are stored in the HDFS directory of the standby cluster.

If you select this option, set the following parameters:

- **Destination NameService Name**: indicates the NameService name of the standby cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
- **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
- **Destination Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
- **Destination Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
- **Destination NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the standby cluster.
- **Target Path**: indicates the HDFS directory for storing standby cluster backup data. The storage path cannot be an HDFS hidden directory, such as a snapshot or recycle bin directory, or a default system directory, such as **/hbase** or **/user/hbase/backup**.
- **Maximum Number of Backup Copies**: indicates the number of backup file sets that can be retained in the backup directory.

Step 8 Set **Backup Content** to one or multiple collections to be backed up.

You can select backup data using either of the following methods:

- Adding a backup data file

a. Click **Add**.
b. Select the table to be backed up under **File Directory**, and click **Add** to add the table to **Backup Content**.
c. Click **OK**.

- Selecting using regular expressions
  a. Click **Query Regular Expression**.
  b. Enter a slash (/) in the first text box. This root directory is not an actual directory but an internal Solr directory.
  c. Enter a regular expression in the second text box. Standard regular expressions are supported. For example, to get indexes containing **solr**, enter **.*solr.***. To get indexes starting with **solr**, enter **solr.***. To get indexes ending with **solr**, enter **.*solr**.
  d. Click **Refresh** to view the displayed tables in **Directory Name**.
  e. Click **Synchronize** to save the result.

  NOTE

  - When entering regular expressions, click + or − to add or delete an expression.
  - If the selected table or directory is incorrect, click **Clear Selected Node** to deselect it.

Step 9 Click **Verify** to check whether the backup task is configured correctly.

The possible causes of the verification failure are as follows:

- The destination active or standby NameNode IP address or NameService name is incorrect.
- The name of the index to be backed up does not exist in the cluster.

Step 10 Click **OK**.

Step 11 In the **Operation** column of the created task in the backup task list, click **More** and select **Back Up Now** to execute the backup task.

After the backup task is executed, the system automatically creates a subdirectory for each backup task in the backup directory. The format of the subdirectory name is *Backup task name_Data source_Task creation time*, and the subdirectory is used to save latest data source backup files. Each time a backup task is executed, a snapshot directory named *_Snapshot absolute seconds* is created in the directory. When the number of snapshot directories is greater than the value of **Maximum Number of Backup Copies**, the earliest directory is automatically deleted.

----End

2.2.2.4 Backing Up OMS Data to a Third-Party Server Outside a Cluster

Scenario

MRS clusters in security mode support backing up their OMS data (including OMS, LDAP, and DBService data) to third-party servers outside them, significantly improving system reliability.

The data is backed up to third-party servers over Secure File Transfer Protocol (SFTP). To use this function, prepare one or more third-party Linux servers (data

can be stored among the servers if multiple servers are configured), install and configure the servers, and execute the backup transmission script on one of these servers. After the backup is successful, the OMS data is stored in a specified backup path. To restore the data, copy the data to a directory on the active management node.

NOTE

This section applies only to physical machine clusters.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Hosts**, click **Add**, and add three third-party servers to the cluster in sequence as prompted.

Step 3 Install the cluster client in a directory, for example, **/opt/hadoopclient**, on the third-party servers.

Step 4 Log in to a third-party server as user **root**.

NOTE

To back up data to one or multiple servers, you need to perform operations only on one third-party server. After scripts are enabled, the system automatically transmits data to each server based on configurations.

Step 5 Run the following command to enter the directory of the backup transmission script:

cd ${BIGDATA_HOME}/om-agent/nodeagent/tools

Step 6 Run the following command to invoke the backup transmission script:

./backupAndTransfer.sh *FusionInsight client installation directory Backup mode Target directory of backup data* **om**

- Two backup modes:
  - **full**: full backup
  - **inc**: incremental backup

For example, to back up OMS data in full backup mode to a third-party server, run the following command:

./backupAndTransfer.sh /opt/hadoopclient full /opt/backup/result om

NOTE

- OMS backup files consist of OMS, LDAP, DBService, and HDFS-hacluster-fsimage backup files.
- User **omm** must have the write permission on the target directory of backup data.

The command output is as follows:

```
Data backup succeeded.
```

----End

2.3 Restoration

2.3.1 Restoring Data on the Management and Control Plane

Restoring Data Using a Manually Backed Up Database File

Step 1 Prepare the backup SQL file in advance and upload it to the active MRS-DB node. For details, see **Manually Backing Up the MRS Database Data**.

Step 2 Log in to the MRS-DB active node as user **opsadmin**, and run the **su - root** command to switch to user **root**.

NOTE

For details about how to check the IP address, active/standby status, and floating IP address of an MRS-DB node, see **Determining the Active/Standby Status of MRS-DB Nodes**.

Step 3 Run the following command to connect to the MRS-DB database:

mysql -P7306 -h*MRS-DB floating IP address* **-udatasightemr -p***Password of database user datasightemr*

Contact O&M engineers to obtain the default password of database user **datasightemr**.

Step 4 Run the following commands to restore data:

use *Database name*;

Example:

use datasightemr;

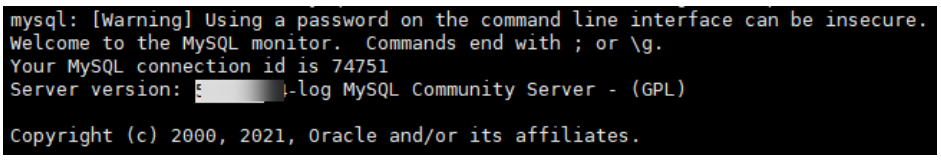
source */tmp/mrs_datasightemr_backup.sql*;

----End

Restoring Data Using a Database File Backed Up on ManageOne

NOTE

The following data restoration procedure applies to newly deployed Huawei Cloud Stack 8.3.0 and later versions (the MySQL version of MRS-DB is 5.7.*x*).



If the current environment is upgraded from an earlier version, perform operations by referring to **Restoring Data Using a Database File Backed Up on ManageOne (Earlier Version)**.

Step 1 Ensure that the S-MRS-MySQL database backup task has been configured using the unified backup function of ManageOne and the task is successfully executed.

Step 2 Obtain the backup file and upload it to the MRS-DB nodes.

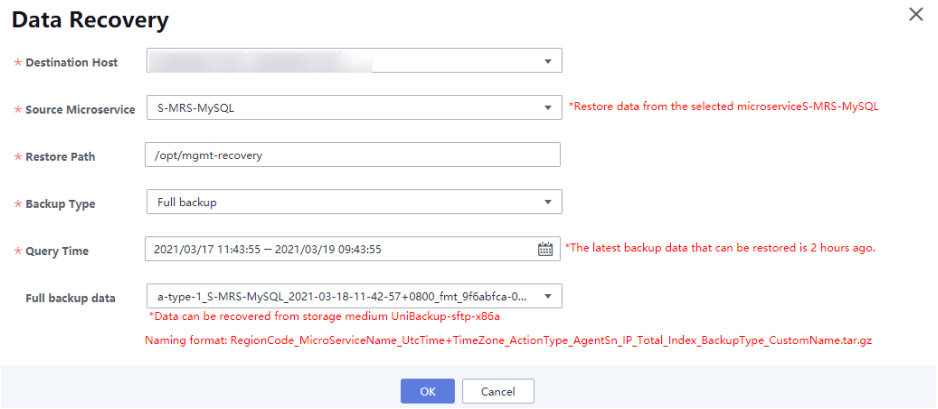
- Obtaining the backup file through the SFTP server
  a. Use WinSCP to log in to the SFTP backup server based on the backup path, username, and password of the SFTP server configured in the backup task on ManageOne.
  b. Go to the backup path of the S-MRS-MySQL database backup task, query the latest backup data based on the backup type and backup time, and download the data to the local host.

     For example, if the backup path is **/backup/*XXX*/CloudDB/S-MRS-MySQL/full**, obtain the backup file **aa-type-1_S-MRS-MySQL_2021-03-17-15-14-14+0800_fmt_9f6abfca-0500-4fee-9348-fcb1cfe9a03e_10.68.6.69_1_1_full_host-24-68_24.tar.gz**.
  c. Upload the obtained backup file to the MRS-DB nodes as user **opsadmin**, for example, to the **/opt/mgmt-recovery** directory.

     NOTE

     To facilitate subsequent operations, you are advised to upload the backup file to both the active and standby MRS-DB nodes.
- Copying the backup file to the MRS-DB nodes on CloudScope
  a. Open a browser, enter https://*CloudScopeLite domain name* in the address box, and press **Enter** to log in to CloudScope

     Log in to CloudScope as a user who has the permission to access CloudBCM.
  b. On the **Services** tab, click **CloudBCM**.
  c. In the navigation tree, choose **Backup Mgmt** > **Backup Center**. In **Cloud Service View**, choose **Cloud** > **CloudDB** > **EI-DB** > **S-MRS-MySQL**.
  d. On the **Backup and Restore** tab, verify that the access status of S-MRS-MySQL is **Accessed** instead of **Backup**. Select the backup task that has been successfully executed and click **Data Recovery**. In the displayed dialog box, click **OK**.
     - In the **Destination Host** drop-down list, select the IP addresses of the associated MRS-DB nodes.
     - Set **Restore Path** to the restoration path of the backup file. Use the default value, for example, **/opt/mgmt-recovery**.
     - Query the latest backup data based on the backup type and backup time. Select **Full backup**, and then select the latest backup file from the **Full backup data** drop-down list.



NOTE

If the backup type of the data to be restored is incremental backup, select **Incr backup** for **Backup Type** and select the backup files generated by the corresponding backup tasks for **Full backup data** and **Incr backup data**.

e. Click **OK**. After the task is successfully executed, the system automatically copies the backup file to the MRS-DB nodes.

Step 3 Log in to the MRS-DB active node as user **opsadmin**, and run the **su - root** command to switch to user **root**.

NOTE

See **Determining the Active/Standby Status of MRS-DB Nodes** for details about how to check the active/standby status of the MRS-DB node and the floating IP address of the database.

Step 4 Log in to the active MRS-DB node as user **opsadmin**, run the **su - root** command to switch to user **root**, and run the following command to restore the database file:

- The following shows an example of full restoration:

  python /data/dbmha/backup_restore/CloudDB_restore.py -p *7306* **-f** */opt/mgmt-recovery/XXX.tar.gz*

  – **-f**: Mandatory. Compressed database backup file.

  – **-p**: specifies the port number of the temporary database instance. Use an idle port, for example, **7306**.

- The following shows an example of incremental restoration:

  Use the **-i** parameter to specify the incremental backup data for restoration.

  Assume that **/opt/mgmt-recovery/YYY.tar.gz** is the incremental backup data file (based on the full backup data file **/opt/mgmt-recovery/XXX.tar.gz**).

  python /data/dbmha/backup_restore/CloudDB_restore.py -p *7306* **-f** */opt/mgmt-recovery/XXX.tar.gz* **-i** */opt/mgmt-recovery/YYY.tar.gz*

The following error information may be displayed during the restoration. However, the database created using the temporary port has been created. Ignore this information.

```
Begin restore,please wait!
extract full backup started ...
extract full backup tar to:
/data/mysql/host-10-28-_10.28.1.70_20220318195346_7306
write my.cnf to:
```



```
/data/mysql/host-10-28-_10.28.1.70_20220318195346_7306/my.cnf
starting mysqld, cmd: su - mysql -c 'nohup /data/mysql/base/bin/mysqld_safe  --defaults-file=/data/mysql/
host-10-28-_10.28.1.70_20220318195346_7306/my.cnf >/dev/null 2>&1 & '
checking mysql by select ...
pid use port[7306] not found
Warning: loop 0, mysql socket not found in port 7306
pid use port[7306] not found
...
Restore database is failure, select check failed, please check it.
```

Record the value of **--defaults-file** in the command output and run the following command:

setsid /data/mysql/base/bin/mysqld_safe --defaults-file=*/data/mysql/host-10-28-_10.28.1.70_20220318195346_7306/my.cnf* **--plugin-dir=/data/mysql/base//lib/plugin --user=mysql**

The command execution takes a long time. Please wait.

After the command is executed, run the following command to check whether you can log in to the temporary database instance:

mysql -P7306 -h*IP address of the current MRS-DB node* **-udatasightemr -p***Password of user datasightemr*

Contact O&M engineers to obtain the default password of database user **datasightemr**.

Step 5 After the temporary database is created, run the following command to export the SQL file:

su - mysql

/data/mysql/base/bin/mysqldump -udatasightemr -h*IP address of the current MRS-DB node* **-P7306 -p***Password of database user datasightemr* **--flush-logs --master-data=2 --single-transaction --set-gtid-purged=OFF --databases mrs datasightemr >** */backup/mysqldump_0312.sql*

Contact O&M engineers to obtain the default password of database user **datasightemr**.

Step 6 Log in to the current active MRS-DB node as user **opsadmin**, run the **su - root** command to switch to user **root**, and run the following commands to restore the database file:

mysql -P7306 -uroot -p*Password of user root* **-S /data/mysql/tmp/mysql.sock**

To obtain the default password of user **root**, see ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator.

source */backup/mysqldump_0312.sql;*

Step 7 After the data is restored, run the following command to stop the temporary database:

ps -ef | grep *mysql_port* **| grep -v grep | awk '{ print "kill -9 " $2 }' | sh**

- **mysql_port**: port number of the temporary database instance.

----End

Restoring Data Using a Database File Backed Up on ManageOne (Earlier Version)

Step 1 Ensure that the S-MRS-MySQL database backup task has been configured using the unified backup function of ManageOne and the task is successfully executed.

Step 2 Obtain the backup file and upload it to the MRS-DB nodes.

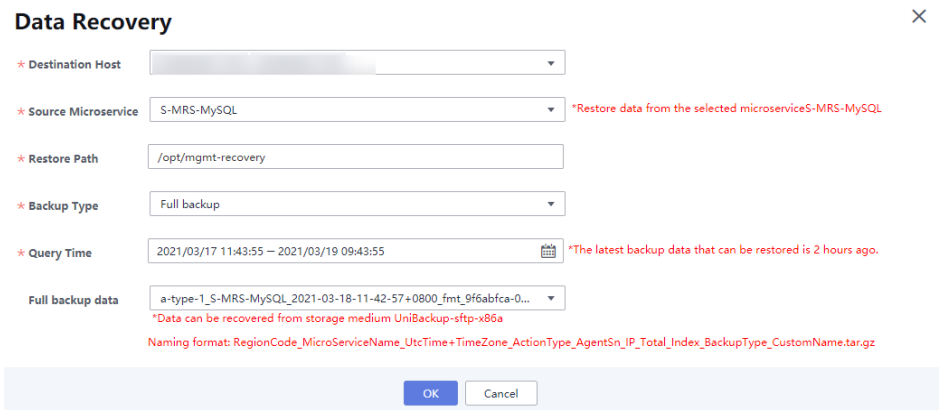
- Obtaining the backup file through the SFTP server
  a. Use WinSCP to log in to the SFTP backup server based on the backup path, username, and password of the SFTP server configured in the backup task on ManageOne.
  b. Go to the backup path of the S-MRS-MySQL database backup task, query the latest backup data based on the backup type and backup time, and download the data to the local host.

     For example, if the backup path is **/backup/*XXX*/CloudDB/S-MRS-MySQL/full**, obtain the backup file **aa-type-1_S-MRS-MySQL_2021-03-17-15-14-14+0800_fmt_9f6abfca-0500-4fee-9348-fcb1cfe9a03e_10.68.6.69_1_1_full_host-24-68_24.tar.gz**.
  c. Upload the obtained backup file to the MRS-DB nodes as user **opsadmin**, for example, to the **/opt/mgmt-recovery** directory.

     NOTE

     To facilitate subsequent operations, you are advised to upload the backup file to both the active and standby MRS-DB nodes.
- Copying the backup file to the MRS-DB nodes on CloudScope
  a. Open a browser, enter https://*CloudScopeLite domain name* in the address box, and press **Enter** to log in to CloudScope

     Log in to CloudScope as a user who has the permission to access CloudBCM.
  b. On the **Services** tab, click **CloudBCM**.
  c. In the navigation tree, choose **Backup Mgmt** > **Backup Center**. In **Cloud Service View**, choose **Cloud** > **CloudDB** > **EI-DB** > **S-MRS-MySQL**.
  d. On the **Backup and Restore** tab, verify that the access status of S-MRS-MySQL is **Accessed** instead of **Backup**. Select the backup task that has been successfully executed and click **Data Recovery**. In the displayed dialog box, click **OK**.
     - In the **Destination Host** drop-down list, select the IP addresses of the associated MRS-DB nodes.
     - Set **Restore Path** to the restoration path of the backup file. Use the default value, for example, **/opt/mgmt-recovery**.
     - Query the latest backup data based on the backup type and backup time. Select **Full backup**, and then select the latest backup file from the **Full backup data** drop-down list.



NOTE

If the backup type of the data to be restored is incremental backup, select **Incr backup** for **Backup Type** and select the backup files generated by the corresponding backup tasks for **Full backup data** and **Incr backup data**.

e. Click **OK**. After the task is successfully executed, the system automatically copies the backup file to the MRS-DB nodes.

Step 3 Log in to the MRS-DB active node as user **opsadmin**, and run the **su - root** command to switch to user **root**.

NOTE

See **Determining the Active/Standby Status of MRS-DB Nodes** for details about how to check the active/standby status of the MRS-DB node and the floating IP address of the database.

Step 4 Run the following commands to manually trigger an active/standby database switchover:

su - mysql

python /usr/local/bin/MHA_SwitchOverManual.py

After the active/standby switchover is successful, connect to the database of the standby MRS-DB node. The current node to which you have logged in is the standby node. Disable the active/standby synchronization to reserve the database before the restoration for rollback.

mysql -P7306 -uroot -p*Password of user root* **-S /data/mysql/tmp/mysql.sock**

stop slave;

To obtain the default password of user **root**, see ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator.

Step 5 Log in to the MRS-DB active node as user **opsadmin**, run the **su - root** command to switch to user **root**, and run the following command to restore the database file:

- The following shows an example of full restoration:

  python /usr/local/CloudAgent/plugins/CloudBasicComponentMgmt/clouddb_backup/CloudDB_restore_mysql.py -p *7306* **-f** */opt/mgmt-recovery/XXX.tar.gz*

  – **-f** (mandatory): specifies the full path of the full backup package.

  - **-p**: specifies the port number of the temporary database instance. Use an idle port, for example, **7306**.

- The following shows an example of incremental restoration:

  Use the **-i** parameter to specify the incremental backup data for restoration.

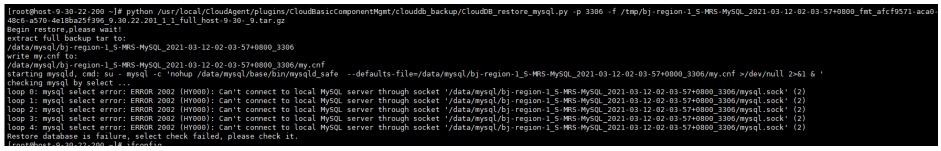
  Assume that **/opt/mgmt-recovery/YYY.tar.gz** is the incremental backup data file (based on the full backup data file **/opt/mgmt-recovery/XXX.tar.gz**).

  python /usr/local/CloudAgent/plugins/CloudBasicComponentMgmt/clouddb_backup/CloudDB_restore_mysql.py -p *7306* **-f** */opt/mgmt-recovery/XXX.tar.gz* **-i** */opt/mgmt-recovery/YYY.tar.gz*

  If there are multiple incremental data packages, use commas (,) to separate them. For example:

  python /usr/local/CloudAgent/plugins/CloudBasicComponentMgmt/clouddb_backup/CloudDB_restore_mysql.py -p *7306* **-f** */opt/mgmt-recovery/XXX.tar.gz* **-i** */opt/mgmt-recovery/YYY.tar.gz,/opt/mgmt-recovery/ZZZ.tar.gz*

If the error shown in the following figure is reported during full or incremental data restoration but you can log in to the database created using the temporary port (7306), the temporary database instance is successfully generated. Ignore the error.



You can run the following command to check whether you can log in to the temporary database instance:

mysql -P7306 -h*IP address of the current MRS-DB node* **-udatasightemr -p***Password of database user datasightemr*

Contact O&M engineers to obtain the default password of database user **datasightemr**.

Step 6 After the temporary database is created, run the following command to export the SQL file:

su - mysql

/data/mysql/base/bin/mysqldump -udatasightemr -h*IP address of the current MRS-DB node* **-P7306 -p***Password of database user datasightemr* **--flush-logs --master-data=2 --single-transaction --set-gtid-purged=OFF --databases mrs datasightemr >** */backup/mysqldump_0312.sql*

Contact O&M engineers to obtain the default password of database user **datasightemr**.

Step 7 Log in to the current active MRS-DB node as user **opsadmin**, run the **su - root** command to switch to user **root**, and run the following commands to restore the database file:

mysql -P7306 -uroot -p*Password of user root* **-S /data/mysql/tmp/mysql.sock**

To obtain the default password of user **root**, see ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator.

source */backup/mysqldump_0312.sql;*

Step 8 After the data is restored, log in to the current MRS-DB standby node as user **opsadmin**, run the **su - root** command to switch to user **root**, and run the following commands to restore the database file and restart the active/standby synchronization:

mysql -P7306 -uroot -p*Password of user root* **-S /data/mysql/tmp/mysql.sock**

start slave;

----End

2.3.2 Restoring Data on the Tenant Plane

2.3.2.1 Restoring OMS Data

Scenario

Manager data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in FusionInsight Manager, an exception occurs or the operation has not achieved the expected result. All modules are faulty and become unavailable.

System administrators can create a restoration task in FusionInsight Manager to recover Manager data. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that of data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the Manager data that is generated after the data backup and before the data restoration will be lost.

Impact on the System

- In the restoration process, the Controller needs to be restarted and FusionInsight Manager cannot be logged in or operated during the restart.
- During the restoration, the cluster needs to be restarted and cannot be accessed during the restart.
- After data restoration, the data, such as system configuration, user information, alarm information, and audit information, that is generated after the data backup and before the data restoration will be lost. This may result in data query failure or cluster access failure.
- After the Manager data is recovered, the system forces the LdapServer of each cluster to synchronize data from the OLadp.

Prerequisites

- To restore data from a remote HDFS, you need to prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, system mutual trust needs to be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The status of the OMS resources and the LdapServer instances of each cluster is normal. If the status is abnormal, data restoration cannot be performed.
- The status of the cluster hosts and services is normal. If the status is abnormal, data restoration cannot be performed.
- The cluster host topologies during data restoration and data backup are the same. If the topologies are different, data restoration cannot be performed and you need to back up data again.
- The services added to the cluster during data restoration and data backup are the same. If the topologies are different, data restoration cannot be performed and you need to back up data again.
- The upper-layer applications that depend on the cluster are stopped.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view the historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restore** > **Restoring Management**. On the displayed page, click **Create**.

Step 4 Set **Task Name** to the name of the restoration task.

Step 5 Set **Recovery Object** to **OMS**.

Step 6 In the **Restoration Configuration** area, select **OMS**.

Step 7 Set **Path Type** of **OMS** to a backup directory type.

The settings vary according to backup directory types:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node.

  If you select **LocalDir**, you also need to set **Source Path** to select the backup file to be restored, for example, *Version_Data source_Task execution time*.**tar.gz**.

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select **LocalHDFS**, set the following parameters:

  - **Source Path**: indicates the full path of the backup file in the HDFS, for example, *Backup path/Backup task name_Task creation time/Version_Data source_Task execution time*.**tar.gz**.
  - **Cluster for Restoration**: Enter the name of the cluster used during restoration task execution.
  - **Source NameService Name**: indicates the NameService name that corresponds to the backup directory when a restoration task is executed. The default value is **hacluster**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select **RemoteHDFS**, set the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time*.**tar.gz**.
  - **Source Cluster**: Select the cluster of the Yarn queue used by the recovery data.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol. If you select **NFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the complete path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time*.**tar.gz**.

- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol. If you select **CIFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select **SFTP**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **OBS**: indicates that backup files are stored in OBS.

  If you select **OBS**, set the following parameters:
  - **Source Path**: indicates the full OBS path of a backup file, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

Step 8 Click **OK**.

Step 9 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

Step 10 Log in to the active and standby management nodes as user **omm**.

Step 11 Run the following command to restart OMS:

sh ${BIGDATA_HOME}/om-server/om/sbin/restart-oms.sh

The command is run successfully if the following information is displayed:

```
start HA successfully.
```

Run **sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh** to check whether **HAAllResOK** of the management node is **Normal** and whether FusionInsight Manager can be logged in again. If yes, OMS is restarted successfully.

Step 12 On FusionInsight Manager, choose **Cluster** > **Services** > **KrbServer**. Click **More** and select **Synchronize Configurations**. In the dialog box displayed, click **OK**. Wait until the KrbServer configuration synchronization is complete.

Step 13 In the upper right corner of **Homepage**, click **More** and select **Synchronize Configurations**. In the dialog box displayed, click **OK**. Wait until the cluster configuration is successfully synchronized.

Step 14 In the upper right corner of **Homepage**, click **More** and select **Restart**. In the dialog box displayed, enter the password of the current login user and click **OK**. Wait until the cluster is successfully restarted.

----End

2.3.2.2 Restoring Metadata and Other Data

2.3.2.2.1 Restoring CDL Data

Scenario

CDL data needs to be restored in the following scenarios: Data is modified or deleted unexpectedly and needs to be restored. After an administrator performs major operations (such as upgrade and data adjustment) on CDL, an exception occurs or the expected result is not achieved. All modules are faulty and become unavailable. Data is migrated to a new cluster.

CDL metadata is stored in DBService and Kafka. A system administrator can create DBService and Kafka restoration tasks on FusionInsight Manager to restore CDL data. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To restore the data when services are normal, manually back up the latest management data first and then restore the data. Otherwise, the DBService and Kafka data that is generated after the data backup and before the data restoration will be lost.
- By default, MRS clusters use DBService to store metadata of Hive, Hue, Loader, Spark, CDL, Redis, and Oozie. Restoring DBService data will restore the metadata of all these components.

Impact on the System

- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is restored, the configurations of the components that depend on DBService may expire and these components need to be restarted.
- After the metadata is restored, the offset information stored on ZooKeeper by Kafka consumers is rolled back, resulting in repeated consumption.

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The status of the active and standby DBService instances is normal. If the status is abnormal, data restoration cannot be performed.
- The Kafka service is disabled first, and then enabled upon data restoration.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the row where the specified backup task is located, choose **More** > **View History** in the **Operation** column to display the historical execution records of the backup task.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.
- **Backup Path**: indicates the full path where backup files are stored.

  Select the correct path, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In the **Restoration Configuration** area, select **DBService** and **Kafka**.

NOTE

If multiple DBService or Kafka services are installed, select the DBService or Kafka services to be restored.

Step 8 Set **Path Type** of **DBService** to a backup directory type. For details about how to configure the parameters, see **Step 8**.

Step 9 Set **Path Type** of **Kafka** to a backup directory type. For details about how to configure the parameters, see **Step 8**.

Step 10 Click **OK**.

Step 11 In the restoration task list, locate the row where the created task is located, and click **Start** in the **Operation** column.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

----End

2.3.2.2.2 Restoring ClickHouse Metadata

Scenario

ClickHouse metadata needs to be restored in the following scenarios: Data is modified or deleted unexpectedly and needs to be restored. After a user performs major operations (such as upgrade and migration) on ClickHouse, an exception occurs or the expected result is not achieved. The ClickHouse component is faulty and becomes unavailable. Data is migrated to a new cluster.

Users can create a ClickHouse restoration task on FusionInsight Manager. Only manual restoration tasks are supported.

NOTICE

- Data can be restored only when the system version during data backup is the same as the current system version.
- To restore ClickHouse metadata when the service is running properly, you are advised to manually back up the latest ClickHouse metadata before restoration. Otherwise, the ClickHouse metadata that is generated after the data backup and before the data restoration will be lost.
- Before ClickHouse metadata restoration, ensure that all ClickHouseServer nodes are running properly. Otherwise, the backup may fail.

Impact on the System

- After the metadata is restored, the data generated after the data backup and before the data restoration is lost.
- After the metadata is restored, the ClickHouse upper-layer applications need to be started.

Prerequisites

- You have checked the path for storing ClickHouse metadata backup files.
- You have prepared a standby cluster if you need to restore data remotely from HDFS. If the active and standby clusters are deployed in security mode and they are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active and standby clusters are deployed in normal mode, no mutual trust is required.
- In the active/standby cluster, when restoring data from the remote HDFS to the local host, ensure that the value of **HADOOP_RPC_PROTECTION** of ClickHouse is the same as that of **hadoop.rpc.protection** of HDFS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of the specified task in the task list, choose **More** > **View History**.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.
- **Backup Path**: indicates the full path for storing backup files.

  Select the correct path and copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **ClickHouse** under **Metadata and other data**.

Step 8 Set **Path Type** of **ClickHouse** to a restoration directory type.

Table 2-27 Path for data restoration

Directory Type	Description
LocalDir	Indicates that data is restored from the local disk of the active management node. If you select this option, you also need to configure the following parameters: • **Source Path**: Enter the name of the backup file to be restored. To obtain the file name, log in to the active OMS node, go to the backup path copied in **Step 2**, and record the name of the metadata package, for example, *Backup task name_Data source_Task execution time***.tar.gz**. • **Logical Cluster**: Enter the ClickHouse logical cluster whose data has been backed up.
RemoteHDFS	Indicates that data is restored from the HDFS directory of the standby cluster. If you select this value option, you also need to configure the following parameters: • **Source NameService Name**: indicates the NameService name of the backup data cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster. • **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**. • **Source Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster. • **Source Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster. • **Source NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the destination cluster. • **Source Path**: Enter the complete HDFS path for storing backup data of the standby cluster, that is, the backup path copied in **Step 2**, for example, *Backup path*/*Backup task name_Data source_Task creation time*. • **Logical Cluster**: Enter the ClickHouse logical cluster whose data has been backed up.

Directory Type	Description
OBS	Indicates that data is restored from OBS. If you select this option, you also need to configure the following parameters: • **Source Path**: indicates the full OBS path of a backup file, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time*.**tar.gz**. • **Logical Cluster**: Enter the ClickHouse logical cluster whose data has been backed up.

The configurations vary based on backup directory types:

Step 9 Click **OK**.

Step 10 In the restoration task list, locate the row where the created task is, and click **Start** in the **Operation** column. In the dialog box that is displayed, click **OK** to start the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be re-executed.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to re-execute the task.

Step 11 Choose **Cluster** > **Services** and start the ClickHouse service.

----End

2.3.2.2.3 Restoring Containers Metadata

Scenario

Restore Containers metadata in the following scenarios: Data is modified or deleted accidentally, or needs to be recovered; data exceptions occur or the change results are not as expected after major operations such as upgrade or data migration are performed on Containers; all modules are faulty and become unavailable; data is migrated to a new cluster.

You can create a restoration task on FusionInsight Manager to restore Containers metadata. Only manual restoration tasks are supported.

NOTICE

- Data can be restored only when the system version during data backup is the same as the current system version.
- To restore Containers metadata when services are running properly, manually back up the latest Containers metadata before restoration. Otherwise, the Containers metadata generated after the data backup and before the data restoration will be lost.
- You are advised to restore the metadata of Containers and RTDService at the same time. If only the RTDService metadata is backed up and restored, RTDService needs to be brought offline and then online. If only the Containers metadata is backed up and restored, the service management status may be inconsistent with the running status.
- After the Containers service is restored, if the prediction variables, model variables, and decision engines of the corresponding event sources have been brought online, you need to manually bring them online again after the restoration.

Impact on the System

- After the metadata is restored, the data generated after the data backup and before the data restoration is lost.
- After the metadata is restored, the upper-layer applications of Containers need to be restarted.

Prerequisites

- You have checked the path for storing Containers metadata backup files.
- You have prepared a standby cluster if you need to restore data remotely from HDFS. If the active cluster is deployed in security mode (with Kerberos authentication enabled) and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode (with Kerberos authentication disabled), mutual trust is not required.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of the specified task in the task list, click **More** and select **View History**.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.
- **Backup Path**: indicates the full path for storing backup files.

  Select the correct path and copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **Containers** under **Metadata and other data**.

Step 8 Set **Path Type** of **Containers** to a restoration directory type.

The configurations vary based on backup directory types:

- **LocalDir**: indicates that data is restored from the local disk of the active management node.

  If you select this option, you also need to set **Source Path**, which indicates the backup file to be restored, for example, *Backup task name_Data source_Task execution time*.**tar.gz**.

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select this option, you also need to configure the following parameters:

  - **Source Path**: indicates the full path for storing backup files in HDFS, for example, *Backup path/Backup task name_Task creation time/Version_Data source_Task execution time*.**tar.gz**.
  - **Source NameService Name**: indicates the NameService name that corresponds to the backup directory when a restoration task is executed. The default value is **hacluster**.

- **RemoteHDFS**: indicates that data is restored from the HDFS directory of the standby cluster.

  If you select this option, you also need to configure the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
  - **Source Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
  - **Source NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the destination cluster.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Data source_Task execution time*.**tar.gz**.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate the row where the created task is, and click **Start** in the **Operation** column. In the dialog box that is displayed, click **OK** to start the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be re-executed.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to re-execute the task.

----End

2.3.2.2.4 Recovering DBService Data

Scenario

DBService data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in DBService, an exception occurs or the operation has not achieved the expected result. All modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a recovery task in FusionInsight Manager to recover DBService data. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that of data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the DBService data that is generated after the data backup and before the data recovery will be lost.
- By default, MRS clusters use DBService to store metadata of Hive, Hue, Loader, Spark, CDL, Redis, and Oozie. Restoring DBService data will restore the metadata of all these components.

Impact on the System

- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is restored, the configurations of the components that depend on DBService may expire and these components need to be restarted.

Prerequisites

- To restore data from a remote HDFS, you need to prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.

- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The status of the active and standby DBService instances is normal. If the status is abnormal, data restoration cannot be performed.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view the historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In the **Restoration Configuration** area, select **DBService**.

NOTE

If multiple DBServices are installed, select the DBServices to be restored.

Step 8 Set **Path Type** of **DBService** to a backup directory type.

The settings vary according to backup directory types:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node.

  If you select **LocalDir**, you also need to set **Source Path** to select the backup file to be restored, for example, *Version_Data source_Task execution time*.**tar.gz**.
- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select **LocalHDFS**, set the following parameters:

  – **Source Path**: indicates the full path of the backup file in the HDFS, for example, *Backup path/Backup task name_Task creation time/Version_Data source_Task execution time*.**tar.gz**.

  - **Source NameService Name**: indicates the NameService name that corresponds to the backup directory when a restoration task is executed. The default value is **hacluster**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select **RemoteHDFS**, set the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select **NFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol.

  If you select **CIFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.

  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select **SFTP**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **OBS**: indicates that backup files are stored in OBS.

  If you select **OBS**, set the following parameters:

  - **Source Path**: indicates the full OBS path of a backup file, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

----End

2.3.2.2.5 Restoring Flink Metadata

Scenario

Flink metadata needs to be restored in the following scenarios: Data is modified or deleted unexpectedly and needs to be restored. After an administrator performs major operations (such as upgrade and data adjustment) on Flink, an exception occurs or the expected result is not achieved. The Flink component is faulty and becomes unavailable. Data is migrated to a new cluster.

System administrators can create a Flink restoration task on FusionInsight Manager. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To restore Flink metadata when the service is running properly, you are advised to manually back up the latest Flink metadata before restoration. Otherwise, the Flink metadata that is generated after the data backup and before the data restoration will be lost.
- Flink metadata restoration and service data restoration cannot be performed at the same time. Otherwise, service data restoration fails. You are advised to restore service data after metadata restoration is complete.

Impact on the System

- Before restoring the metadata, you need to stop the Flink service. During this period, all upper-layer applications are affected and cannot work properly.
- After the metadata is restored, the data generated after the data backup and before the data restoration is lost.
- After the metadata is restored, the Flink upper-layer applications of Solr need to be started.

Prerequisites

- You have checked the path for storing Flink metadata backup files.
- The Flink service has been stopped before its metadata is restored.
- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of the specified task in the task list, choose **More** > **View History**.

In the displayed window, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct path, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **Flink** under **Metadata and other data**.

Step 8 Set **Path Type** of **Flink** to a restoration directory type.

The settings vary according to backup directory types:

- **LocalDir**: indicates that data is restored from the local disk of the active management node.

  If you select **LocalDir**, you also need to set **Source Path** to select the backup file to be restored, for example, *Backup task name_Data source_Task execution time***.tar.gz**.

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select **LocalHDFS**, set the following parameters:

  - **Source Path**: indicates the full path of the backup file in the HDFS, for example, *Backup path/Backup task name_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Source NameService Name**: indicates the NameService name that corresponds to the backup directory when a restoration task is executed.

- **RemoteHDFS**: indicates that data is restored from the HDFS directory of the standby cluster.

  If you select **RemoteHDFS**, set the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Data source_Task execution time***.tar.gz**.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the source cluster.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate the row where the created task is located, and click **Start** in the **Operation** column. In the displayed dialog box, click **OK** to start the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

Step 11 Choose **Cluster** > **Services** and start the Flink service.

----End

2.3.2.2.6 Recovering HBase Metadata

Scenario

To ensure HBase metadata security (including tableinfo files and HFiles) or before a major operation on HBase system tables (such as upgrade or migration), you need to back up HBase metadata to prevent HBase service unavailability caused by HBase system table directory or file damages. The backup data can be used to recover the system if an exception occurs or the operation has not achieved the expected result, minimizing the adverse impacts on services.

System administrators can create a recovery task in FusionInsight Manager to recover HBase metadata. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the HBase data that is generated after the data backup and before the data recovery will be lost.
- It is recommended that a data restoration task restore the metadata of only one component to prevent the data restoration of other components from being affected by stopping a service or instance. If data of multiple components is restored at the same time, data restoration may fail.

  HBase metadata cannot be restored at the same time as NameNode metadata. As a result, data restoration fails.

Impact on the System

- Before restoring the metadata, you need to stop the HBase service, during which the HBase upper-layer applications are unavailable.
- After the metadata is restored, the data generated after the data backup and before the data restoration is lost.
- After the metadata is restored, the upper-layer applications of HBase need to be started.

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- You have checked the path for storing HBase metadata backup files.
- The HBase service has been stopped before its metadata is restored.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view the historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **HBase** under **Metadata and other data**.

NOTE

If multiple HBase services are installed, select the HBase services to be restored.

Step 8 Set **Path Type** of **HBase** to a backup directory type.

The settings vary according to backup directory types:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node.

  If you select **LocalDir**, you also need to set **Source Path** to select the backup file to be restored, for example, *Version_Data source_Task execution time*.**tar.gz**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select **RemoteHDFS**, set the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.

- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol.

  If you select **NFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

- **CIFS**: indicates that backup files are stored in NAS using the CIFS protocol.

  If you select **CIFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

If you select **SFTP**, set the following parameters:

- **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
- **Server IP Address**: indicates the IP address of the server where the backup data is stored.
- **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
- **Username**: indicates the username for connecting to the server using the SFTP protocol.
- **Password**: indicates the password for connecting to the server using the SFTP protocol.
- **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

- **OBS**: indicates that backup files are stored in OBS.

  If you select **OBS**, set the following parameters:

  - **Source Path**: indicates the full OBS path of a backup file, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

----End

2.3.2.2.7 Restoring IoTDB Metadata

Scenario

To ensure IoTDB metadata security and prevent the IoTDB service from being unavailable due to IoTDB file damage, IoTDB metadata needs to be backed up. In this way, the system can restore data timely when an exception is reported or an operation does not achieve the expected result, minimizing the impact on services.

System administrators can create an IoTDB restoration task on FusionInsight Manager. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To restore the data when services are normal, manually back up the latest management data first and then restore the data. Otherwise, the IoTDB data that is generated after the data backup and before the data restoration will be lost.
- You are advised to restore the metadata of only one component in a restoration task to prevent the stop of a service or instance from affecting the data restoration of other components. If data of multiple components is restored at the same time, data restoration may fail.

Impact on the System

After the metadata is restored, the data generated after the data backup and before the data restoration is lost.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the row where the specified backup task is located, choose **More** > **View History** in the **Operation** column to display the historical execution records of the backup task.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.
- **Backup Path**: indicates the full path where backup files are stored.

  Select the correct path, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **IoTDB** under **Metadata and other data**.

NOTE

If multiple IoTDB services are installed, select the IoTDB service to be restored.

Step 8 Select a backup directory type for **Path Type**.

The configurations vary based on backup directory types:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node.

  If you select this option, you also need to set **Source Path**, which indicates the backup file to be restored, for example, *Version_Data source_Task execution time***.tar.gz**.

- **NFS**: indicates that backup files are stored in NAS using the NFS protocol.

  If you select this option, configure the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the complete path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, configure the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
  - **Source Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
  - **Source NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the standby cluster.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

- **CIFS**: indicates that backup files are stored in NAS using the CIFS protocol.

  If you select this option, configure the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.

  - **Source Path**: indicates the complete path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time*.**tar.gz**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select this option, configure the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the complete path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time*.**tar.gz**.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate the row where the created task is located, and click **Start** in the **Operation** column.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

Step 11 Choose **Cluster** > **Services** and start the IoTDB service.

----End

2.3.2.2.8 Recovering Kafka Metadata

Scenario

Kafka data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in ZooKeeper, an exception occurs or the operation has not achieved the expected result. All Kafka modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a recovery task in FusionInsight Manager to recover Kafka data. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To restore Kafka metadata when the service is running properly, you are advised to manually back up the latest Kafka metadata before restoration. Otherwise, the Kafka metadata that is generated after the data backup and before the data restoration will be lost.
- The content of this section is available for Kafka metadata restoration and not for service data restoration.

Impact on the System

- After the metadata is restored, the data generated after the data backup and before the data restoration is lost.
- After the metadata is restored, the offset information stored on ZooKeeper by Kafka consumers is rolled back, resulting in repeated consumption.

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The Kafka service is disabled first, and then enabled upon data restoration.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In the **Restoration Configuration** area, select **Kafka**.

NOTE

If multiple Kafka services are installed, select the Kafka services to be restored.

Step 8 Set **Path Type** of **Kafka** to a backup directory type.

The settings vary according to backup directory types:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node.

  If you select **LocalDir**, you also need to set **Source Path** to select the backup file to be restored, for example, *Version_Data source_Task execution time*.**tar.gz**.

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select **LocalHDFS**, set the following parameters:

  - **Source Path**: indicates the full path of the backup file in the HDFS, for example, *Backup path/Backup task name_Task creation time/Version_Data source_Task execution time*.**tar.gz**.
  - **Source NameService Name**: indicates the NameService name that corresponds to the backup directory when a restoration task is executed. The default value is **hacluster**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select **RemoteHDFS**, set the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time*.**tar.gz**.

  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
- **NFS**: indicates that backup files are stored in NAS using the NFS protocol.

  If you select **NFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **CIFS**: indicates that backup files are stored in NAS using the CIFS protocol.

  If you select **CIFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **OBS**: indicates that backup files are stored in OBS.

  If you select **OBS**, set the following parameters:
  - **Source Path**: indicates the full OBS path of a backup file, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

NOTICE

- If the Kafka service is reinstalled and metadata is restored after data backup, or metadata is migrated to a new cluster, the Kafka broker cannot be started. View the error in the **/var/log/Bigdata/kafka/broker/server.log** file. An example is as follows:

  ```
  ERROR Fatal error during KafkaServer startup. Prepare to shutdown (kafka.server.KafkaServer)kafka.common.InconsistentClusterIdException: The Cluster ID kVSgfurUQFGGpHMTBqBPiw doesn't match stored clusterId Some(0Qftv9yBTAmf2iDPSlIk7g) in meta.properties. The broker is trying to join the wrong cluster. Configured zookeeper.connect may be wrong. at kafka.server.KafkaServer.startup(KafkaServer.scala:220) at kafka.server.KafkaServerStartable.startup(KafkaServerStartable.scala:44) at kafka.Kafka $.main(Kafka.scala:84) at kafka.Kafka.main(Kafka.scala)
  ```

  Check the value of **log.dirs** in the Kafka Broker configuration file **${BIGDATA_HOME}/FusionInsight_Current/*Broker/etc/server.properties**. The value is the Kafka data directory. Go to the Kafka data directory and change the value **0Qftv9yBTAmf2iDPSlIk7g** of **cluster.id** in **meta.properties** to **kVSgfurUQFGGpHMTBqBPiw** (the latest value in the error log).
- The preceding modification must be performed on each node where Broker is located. After the modification, restart the Kafka service.

----End

2.3.2.2.9 Recovering NameNode Data

Scenario

NameNode data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in NameNode, an exception occurs or the operation has not achieved the expected result. All modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a recovery task in FusionInsight Manager to recover NameNode data. Only manual restoration tasks are supported.

NOTICE

- If HDFS service data also needs to be restored, restore HDFS service data first and then NameNode data.
- Data restoration can be performed only when the system version is consistent with that during data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the NameNode data that is generated after the data backup and before the data recovery will be lost.
- It is recommended that a data restoration task restore the metadata of only one component to prevent the data restoration of other components from being affected by stopping a service or instance. If data of multiple components is restored at the same time, data restoration may fail.

  HBase metadata cannot be restored at the same time as NameNode metadata. As a result, data restoration fails.

Impact on the System

- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is recovered, the NameNode needs to be restarted and is unavailable during the restart.
- After data is restored, metadata and service data may not be matched, the HDFS enters the security mode, and the HDFS service fails to be started. .

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.
- On FusionInsight Manager, all the NameNode role instances whose data is to be recovered are stopped. Other HDFS role instances must keep running. After data is recovered, the NameNode role instances need to be restarted. The NameNode role instances cannot be accessed during the restart.
- The NameNode backup files are stored ***Data path*/LocalBackup/** on the active management node.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **HDFS**. On the displayed page, click **Instance** then **NameNode** to check whether the NameNode instances whose data is to be restored are stopped. If they are not, stop them.

Step 2 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 3 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 4 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 5 Click **Create**.

Step 6 Set **Task Name** to the name of the restoration task.

Step 7 Select the desired cluster from **Recovery Object**.

Step 8 In the **Restoration Configuration** area, select **NameNode**.

Step 9 Set **Path Type** of **NameNode** to a backup directory type.

The settings vary according to backup directory types:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node.

  If you select **LocalDir**, set the following parameters:

  - **Source Path**: indicates the full path of the backup file on the local disk, for example, *Backup path/Backup task name_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select **RemoteHDFS**, set the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.

  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution. The name must be the same as the name of the queue that is running properly in the cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol. If you select **NFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol. If you select **CIFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select **SFTP**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.

  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
- **OBS**: indicates that backup files are stored in OBS.

  If you select **OBS**, set the following parameters:
  - **Source Path**: indicates the full OBS path of a backup file, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.

Step 10 Click **OK**.

Step 11 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

Step 12 On FusionInsight Manager, choose **Cluster** > **Services** > **HDFS**. Click **More** and select **Restart Service**.

On the displayed page, enter the password of the administrator who has logged in for authentication and click **OK**. After the system displays "Operation succeeded", click **Finish**. The service is started successfully.

----End

2.3.2.2.10 Restoring Redis Data

Scenario

To ensure Redis data security or before and after a critical operation (such as upgrade and migration) on Redis, Redis data needs to be backed up. The backup data can be used to recover the system in time if an exception occurs or the expected result has not been achieved, minimizing the adverse impact on services.

System administrators can create a recovery task in FusionInsight Manager to recover Redis data. Only manual restoration tasks are supported.

NOTICE

Data restoration can be performed only when the system version is consistent with that during data backup.

Impact on the System

After the data is restored, the data generated after the data backup and before the data restoration is lost.

Prerequisites

- You have checked that HDFS services have been deployed in the current cluster.
- The name of the cluster created by Redis is the same as the cluster name in the backup task.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In the **Restoration Configuration** area, select **Redis**.

Step 8 Set **Path Type** of **Redis** to a backup directory type.

The settings vary according to backup directory types:

- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select **LocalHDFS**, set the following parameters:

- **Source Path**: indicates the full path of the backup file in the HDFS. The specific path is the backup path set in **Step 2**.
- **Source NameService Name**: indicates the NameService name that corresponds to the backup directory when a restoration task is executed. The default value is **hacluster**.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

----End

2.3.2.2.11 Restoring RTDService Metadata

Scenario

Restore RTDService metadata in the following scenarios: Data is modified or deleted accidentally, or needs to be recovered; data exceptions occur or the change results are not as expected after major operations such as upgrade or data migration are performed on RTDService; all modules are faulty and become unavailable; data is migrated to a new cluster.

You can create a restoration task on FusionInsight Manager to restore RTDService metadata. Only manual restoration tasks are supported.

NOTICE

- Data can be restored only when the system version during data backup is the same as the current system version.
- To restore RTDService metadata when services are running properly, manually back up the latest RTDService metadata before restoration. Otherwise, the RTDService metadata generated after the data backup and before the data restoration will be lost.
- You are advised to restore the metadata of RTDService and Containers at the same time. If only the RTDService metadata is backed up and restored, RTDService needs to be brought offline and then online. If only the Containers metadata is backed up and restored, the service management status may be inconsistent with the running status.

Impact on the System

- After the metadata is restored, the data generated after the data backup and before the data restoration is lost.
- After the metadata is restored, the upper-layer applications of RTDService need to be restarted.

Prerequisites

- You have checked the path for storing RTDService metadata backup files.
- You have prepared a standby cluster if you need to restore data remotely from HDFS. If the active cluster is deployed in security mode (with Kerberos authentication enabled) and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode (with Kerberos authentication disabled), mutual trust is not required.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of the specified task in the task list, click **More** and select **View History**.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.
- **Backup Path**: indicates the full path for storing backup files.

  Select the correct path and copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **RTDService** under **Metadata and other data**.

Step 8 Set **Path Type** of **RTDService** to a restoration directory type.

The configurations vary based on backup directory types:

- **LocalDir**: indicates that data is restored from the local disk of the active management node.

  If you select this option, you also need to set **Source Path**, which indicates the backup file to be restored, for example, *Backup task name_Data source_Task execution time*.**tar.gz**.
- **LocalHDFS**: indicates that the backup files are stored in the HDFS directory of the current cluster.

  If you select this option, you also need to configure the following parameters:

  – **Source Path**: indicates the full path for storing backup files in HDFS, for example, *Backup path/Backup task name_Task creation time/Version_Data source_Task execution time*.**tar.gz**.

  - **Source NameService Name**: indicates the NameService name that corresponds to the backup directory when a restoration task is executed. The default value is **hacluster**.

- **RemoteHDFS**: indicates that data is restored from the HDFS directory of the standby cluster.

  If you select this option, you also need to configure the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
  - **Source Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
  - **Source NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the destination cluster.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time/Data source_Task execution time*.**tar.gz**.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate the row where the created task is, and click **Start** in the **Operation** column. In the dialog box that is displayed, click **OK** to start the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be re-executed.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to re-execute the task.

----End

2.3.2.2.12 Restoring Solr Metadata

Scenario

Solr metadata needs to be restored in the following scenarios: Data is modified or deleted unexpectedly and needs to be restored. After an administrator performs major operations (such as upgrade and data adjustment) on ZooKeeper, an exception occurs or the expected result is not achieved. The Solr component is faulty and becomes unavailable. Data is migrated to a new cluster.

System administrators can create a Solr restoration task on FusionInsight Manager. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To restore Solr metadata when the service is running properly, you are advised to manually back up the latest Solr metadata before restoration. Otherwise, the Solr metadata that is generated after the data backup and before the data restoration will be lost.

Impact on the System

- Before restoring the metadata, you need to stop the Solr service. During this period, all upper-layer applications are affected and cannot work properly.
- After the metadata is restored, the data generated after the data backup and before the data restoration is lost.
- After the metadata is restored, the upper-layer applications of Solr need to be started.

Prerequisites

- If you need to restore data from a remote Solr, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- You have checked the path for storing Solr metadata backup files.
- The Solr service has been stopped before its metadata is restored.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the row where the specified backup task is located, choose **More** > **View History** in the **Operation** column to display the historical execution records of the backup task.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.
- **Backup Path**: indicates the full path where backup files are stored.

  Select the correct path, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **Solr** under **Metadata and other data**.

NOTE

If multiple Solr services are installed, select the Solr service to be restored.

Step 8 Select a backup directory type for **Path Type**.

The configurations vary based on backup directory types:

- **LocalDir**: indicates that the backup files are stored on the local disk of the active management node.

  If you select this option, you also need to set **Source Path**, which indicates the backup file to be restored, for example, *Version_Data source_Task execution time*.**tar.gz**.

- **NFS**: indicates that backup files are stored in NAS using the NFS protocol.

  If you select this option, configure the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the complete path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time*.**tar.gz**.

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select this option, configure the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
  - **Source Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
  - **Source NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the standby cluster.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time*.

- **CIFS**: indicates that backup files are stored in NAS using the CIFS protocol.

  If you select this option, configure the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the complete path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
- **SFTP**: indicates that backup files are stored in the backup server using the SFTP protocol.

  If you select this option, configure the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the complete path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate the row where the created task is located, and click **Start** in the **Operation** column.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

Step 11 Log in to FusionInsight Manager and restart the Solr service.

----End

2.3.2.3 Restoring Service Data

2.3.2.3.1 Restoring ClickHouse Service Data

Scenario

ClickHouse data needs to be restored in the following scenarios: Data is modified or deleted unexpectedly and needs to be restored. After a user performs major operations (such as upgrade and migration) on ClickHouse, an exception occurs or the expected result is not achieved. All modules are faulty and become unavailable. Data is migrated to a new cluster.

You can create a restoration task on FusionInsight Manager to restore ClickHouse data. Only manual restoration tasks are supported.

The ClickHouse backup and restoration functions cannot identify the service and structure relationships of objects such as ClickHouse tables, indexes, and views. When executing backup and restoration tasks, you need to manage unified restoration points based on service scenarios to ensure proper service running.

NOTICE

- Data can be restored only when the system version during data backup is the same as the current system version.
- To restore the data when services are normal, manually back up the latest management data first and then restore the data. Otherwise, the ClickHouse data that is generated after the data backup and before the data restoration will be lost.
- Before ClickHouse service data restoration, ensure that all ClickHouseServer nodes are running properly. Otherwise, the backup may fail.

Impact on the System

- During data restoration, user authentication stops and users cannot create new connections.
- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is restored, the ClickHouse upper-layer applications need to be started.

Prerequisites

- You have prepared a standby cluster if you need to restore data remotely from HDFS. If the active and standby clusters are deployed in security mode and they are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active and standby clusters are deployed in normal mode, no mutual trust is required.
- The time on the active and standby clusters must be the same, and the NTP service on the active and standby clusters uses the same time source.
- The database for storing restored data tables, the location for storing the data tables in HDFS, and the list of users who can access the restored data have been planned.

- The ClickHouse backup file save path is correct.
- The ClickHouse upper-layer applications are stopped.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.
- In the active/standby cluster, when restoring data from the remote HDFS to the local host, ensure that the value of **HADOOP_RPC_PROTECTION** of ClickHouse is the same as that of **hadoop.rpc.protection** of HDFS.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the row where the specified backup task is located, choose **More** > **View History** in the **Operation** column to display the historical execution records of the backup task.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.
- **Backup Path**: indicates the full path for storing backup files.

  Select the correct path and copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **ClickHouse** under **Service data**.

Step 8 Set **Path Type** of **ClickHouse** to a restoration directory type.

Table 2-28 Path for data restoration

Directory Type	Description
RemoteHDFS	• Indicates that the backup files are stored in the HDFS directory of the standby cluster. If you select this option, you also need to configure the following parameters: – **Source NameService Name**: indicates the NameService name of the backup data cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster. – **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**. – **Source Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster. – **Source Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster. – **Source NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the destination cluster. – **Source Path**: indicates the full path of the HDFS directory for storing backup data of the standby cluster. For details, see **Backup Path** obtained in **Step 2**. for example, *Backup path/Backup task name_Data source_Task creation time/*.
OBS	Indicates that data is restored from OBS. If you select this option, you also need to configure the following parameters: • **Source Path**: indicates the full OBS path of a backup file, for example, *Backup path/Backup task name_Data source_Task creation time/Version_Data source_Task execution time*.**tar.gz**.

Step 9 Click **OK**.

Step 10 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be re-executed.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to re-execute the task.

----End

2.3.2.3.2 Restoring Doris Service Data

Scenario

Doris data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in the Doris, an exception occurs or the operation has not achieved the expected result. All modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a recovery task in FusionInsight Manager to recover Doris data. Only manual restoration tasks are supported.

When executing backup and restoration tasks, you need to manage unified restoration points based on service scenarios to ensure proper service running.

NOTICE

- Data can be restored only when the system version during data backup is the same as the current system version.
- To restore data when services are normal, manually back up the latest management data before restoring data. Otherwise, the Doris data generated after data backup and before data restoration will be lost.

Impact on the System

After data is restored, the data generated after data backup and before data restoration is lost.

Prerequisites

- To restore data from a remote HDFS, the following conditions must be met:
  - Prepare a standby cluster for restoring data, and ensure that data in this cluster has been backed up. For details, see **Backing Up Doris Data**. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, you do not need to configure mutual trust.
  - At least one DBroker instance of the Doris service has been deployed in the active cluster.
  - The time on the active and standby clusters must be the same, and the NTP service on the active and standby clusters uses the same time source.
  - The value of **hadoop.rpc.protection** of Doris must be the same as that of **hadoop.rpc.protection** of HDFS in both active and standby clusters.
- If you want to restore data from OBS, you have connected the Doris cluster to OBS and have the permission to access OBS. For details, see "Configuring Interconnection between Doris and OBS" in the MapReduce Service (MRS) 3.3.1-LTS User Guide (for Huawei Cloud Stack 8.3.1) in the MapReduce Service (MRS) 3.3.1-LTS Usage Guide (for Huawei Cloud Stack 8.3.1).

- The database for storing restored data tables, the location for storing the data tables in HDFS, and the list of users who can access the restored data have been planned.
- Check the path for storing Doris backup files.
- Stop the upper-layer Doris applications.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, click **More** and select **View History** to view historical execution records of backup tasks.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.
- **Backup Path**: indicates the full path for storing backup files.

  Select the correct path and copy the full path of backup files in **Backup Path**.

Step 3 Choose **Restoration Management** and click **Create**.

Step 4 Set **Task Name** to the name of the restoration task.

Step 5 Select the desired cluster from **Recovery Object**.

Step 6 In **Restoration Configuration**, select **Doris** under **Service data**.

Step 7 Set **Path Type** of **Doris** to a restoration directory type.

Table 2-29 Path for data restoration

Directory Type	Description
RemoteHDFS	The backup files are stored in the HDFS directory of the standby cluster. If you select this option, you also need to configure the following parameters: • **Source NameService Name**: indicates the NameService name of the backup data cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster. • **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**. • **Source NameNode IP Address**: indicates the service plane IP address of the NameNode in the standby cluster. • **Source NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS configuration of the standby cluster. • **DBroker IP**: indicates the IP address of a service plane where the DBroker role in the cluster is deployed. The DBroker is used to transmit data during restoration. • **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster. For details, see the **Backup Path** obtained in step **Step 2**, for example, *Backup path/Backup task name_Data source_Task creation time/*.
OBS	Data is restored from OBS. If you select this option, you also need to configure the following parameters: **Source Path**: indicates the full OBS directory of the backup files. Specify this path by referring to **Step 2**, for example, *Backup path/Backup task name_Data source_Task creation time/*.

Step 8 Click **Refresh** and select a Doris backup file set that has been backed up.

Step 9 In the **Data Configuration** area, select one or more pieces of backup data for **Select Data** based on service requirements.

Configuration restrictions are as follows:

- There is a database with the same name as the original database of the selected backup data in the Doris of the cluster.
- The backup data is restored to the backup table with the same name as the original table in the database.

- If there is a table with the same name in Doris, ensure that the structures of the two tables are the same, including table names, columns, partitions, and materialized views.

Step 10 Set **Original Configurations** to **true**, indicating that the configuration of the backup data, such as the number of copies, will be used. If this parameter is set to **false**, the default configuration is used to create a table.

Step 11 Click **OK**.

Step 12 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be re-executed.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to re-execute the task.

----End

2.3.2.3.3 Restoring Elasticsearch Service Data

Scenario

Elasticsearch service data needs to be recovered in the following scenarios: Data is modified or deleted unexpectedly and needs to be restored; after an administrator performs a critical operation (such as upgrade or critical data adjustment) on Elasticsearch, an exception occurs or the operation has not achieved the expected result, causing all modules to be faulty; data is migrated to a new cluster.

This section describes how to create an Elasticsearch service data restoration task on FusionInsight Manager. The data can only be manually restored.

NOTICE

- Data can be restored only when the system version during data backup is the same as the current system version.
- To restore Elasticsearch service data when services are normal, manually back up the latest service data first and then restore the service data. Otherwise, the Elasticsearch service data that is generated after the data backup and before the data recovery will be lost.
- The number of index shards to be restored must be the same as the number of index shards in the snapshot.
- During the restoration task execution, the indexes to be restored are automatically closed. After the restoration task is complete, the indexes are automatically opened. If the indexes to be restored do not exist, the indexes are automatically created. Therefore, service operations of the indexes may be affected during the restoration task execution.
- The Elasticsearch service data restoration needs to invoke the snapshot interface through the EsNode1 instance. Therefore, ensure that all EsNode1 instances in the cluster are in good health status and can receive requests normally. To ensure successful restoration, do not perform operations such as adding, deleting, stopping, or restarting Elasticsearch instances, stopping or restarting the Elasticsearch service, or stopping or restarting the cluster.

Impact on the System

- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is recovered, the Elasticsearch upper-layer applications need to be started.

Prerequisites

- The directory for storing the Elasticsearch backup files has been checked.
- The Elasticsearch upper-layer applications have been stopped.
- You have prepared a standby cluster if you need to restore data remotely from HDFS. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- The time on the active and standby clusters must be the same, and the NTP service on the active and standby clusters uses the same time source.

Procedure

Step 1 Log in to FusionInsight Manager, and choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of the specified task in the task list, choose **More** > **View History**.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.
- **Backup Path**: indicates the full path for storing backup files.

  Select the correct path and copy the full path of backup files in **Backup Path**.

Step 3 Choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In the **Restoration Configuration** area, select **Elasticsearch** under **Service data**.

Step 8 In the displayed **Elasticsearch** area, set **Path Type** to the restoration directory type.

The following restoration directory types are supported:

- **RemoteHDFS**: indicates that data is restored from the HDFS directory of the standby cluster.

  If you select this option, you also need to configure the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
  - **Source Path**: indicates the full path for storing backup files in HDFS, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Restoration Point List**: Click **Refresh** and select an Elasticsearch snapshot that has been backed up in the standby cluster.
- **NFS**: indicates that backup files are obtained from the NAS through the NFS protocol. If you select this option, you also need to configure the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Restoration Point List**: Click **Refresh** and select an Elasticsearch snapshot that has been backed up in the NAS.

Step 9 In the **Data Configuration** area, select one or more pieces of backed up data for **Backup Object** based on service requirements.

NOTE

If the **.security_info** or **.index_owner_info** index is selected, disable it in advance by referring to "Running curl Commands in Linux" > "Disabling Indexes" in *MapReduce Service (MRS) 3.3.1-LTS User Guide (for Huawei Cloud Stack 8.3.1) in the MapReduce Service (MRS) 3.3.1-LTS Usage Guide (for Huawei Cloud Stack 8.3.1)*. Otherwise, the restoration task may fail.

Step 10 Specify **Force recovery**. The value **false** does not take effect. All backup data is forcibly restored when there are indexes with the same name. If the index contains data added after the backup, the new data will be lost after the data restoration.

Step 11 Click **Verify** to check whether the restoration task is configured correctly.

- If the specified directory to be recovered does not exist, the verification fails.
- If the forcible replacement conditions are not met, the verification fails.

Step 12 Click **OK**.

Step 13 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be re-executed.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to re-execute the task.
- If the restoration type is RemoteHDFS and Elasticsearch has been deleted and then added again, the message "Failed to obtain the remote snapshot" is displayed. To handle this exception, you need to run the snapshot creation command by referring to **How Do I Prepare for Restoring RemoteHDFS Tasks After Elasticsearch Is Reinstalled?** and then click **Start**.

----End

2.3.2.3.4 Recovering HBase Service Data

Scenario

HBase data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in HBase, an exception occurs or the operation has not achieved the expected result. All modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a recovery task in FusionInsight Manager to recover HBase data. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the HBase data that is generated after the data backup and before the data recovery will be lost.
- HBase service data and HBase/HDFS metadata cannot be restored at the same time. Otherwise, metadata restoration fails. Restore service data after metadata restoration is complete.

Impact on the System

- During the data recovery process, the system disables the HBase table to be recovered and the table cannot be accessed in this moment. The data recovery process takes several minutes, during which the HBase upper-layer applications are unavailable.
- During data restoration, user authentication stops and users cannot create new connections.
- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is recovered, the HBase upper-layer applications need to be started.

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The directory for saving the backup file has been checked.
- The HBase upper-layer applications have been stopped.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **HBase** under **Service Data**.

Step 8 Set **Path Type** of **HBase** to a backup directory type.

The following backup directory types are supported:

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster. If you select **RemoteHDFS**, set the following parameters:
  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of the backup file in the HDFS, for example, *Backup path/xxx/Backup task name_Data source_Task creation time*. To obtain the backup path, locate the row that contains the target backup task in the backup management list, click **More** > **View History** in the **Operation** column, and click **View** in the **Backup Path** column.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **NFS**: indicates that backup files are stored in the NAS using the NFS protocol. If you select **NFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/xxx/Backup task name_Data source_Task creation time*. To obtain the backup path, locate the row that contains the target backup task in the backup management list, click **More** > **View History** in the **Operation** column, and click **View** in the **Backup Path** column.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **CIFS**: indicates that backup files are stored in the NAS using the CIFS protocol. If you select **CIFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/xxx/Backup task name_Data source_Task creation time*. To obtain the backup path, locate the row that contains the target backup task in the backup management list, click **More** > **View History** in the **Operation** column, and click **View** in the **Backup Path** column.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select **SFTP**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/xxx/Backup task name_Data source_Task creation time*. To obtain the backup path, locate the row that contains the target backup task in the backup management list, click **More** > **View History** in the **Operation** column, and click **View** in the **Backup Path** column.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **OBS**: indicates that backup files are stored in OBS.

  If you select this option, you also need to configure the following parameters:
  - **Source Path**: indicates the full OBS path of a backup file, for example, *Backup path/xxx/Backup task name_Data source_Task creation time/Version_Data source_Task execution time***.tar.gz**.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an OBS directory that has been backed up.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

Step 9 Set **Backup Data** column in **Data Configuration** to one or multiple backup data sources to be recovered. In the **Target Namespace** column, specify the target naming space after backup data recovery.

You are advised to set **Target Namespace** to a location that is different from the backup naming space.

Step 10 Set **Force recovery** to **true**, which indicates to forcibly recover all backup data when a data table with the same name already exists. If the data table contains new data added after backup, the new data will be lost after the data recovery. If

you set the parameter to **false**, the restoration task is not executed if a data table with the same name exists.

Step 11 Click **Verify** to check whether the restoration task is configured correctly.

- If the queue name is incorrect, the verification fails.
- If the specified naming space does not exist, the verification fails.
- If the forcible overwrite conditions are not met, the verification fails.

Step 12 Click **OK** to save the settings.

Step 13 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

Step 14 Check whether HBase data is restored in an environment where HBase is newly installed or reinstalled.

- If yes, the administrator needs to set new permission for roles on FusionInsight Manager based on the original service plan.
- If no, no further operation is required.

----End

2.3.2.3.5 Recovering HDFS Service Data

Scenario

HDFS data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in the HDFS, an exception occurs or the operation has not achieved the expected result. All modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a recovery task in FusionInsight Manager to recover HDFS data. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the HDFS data that is generated after the data backup and before the data recovery will be lost.
- The HDFS restoration operation cannot be performed for the directories used by running Yarn tasks, for example, **/tmp/logs**, **/tmp/archived**, and **/tmp/hadoop-yarn/staging**. Otherwise, data restoration using Distcp tasks fails due to file loss.

Impact on the System

- During data restoration, user authentication stops and users cannot create new connections.
- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is recovered, the HDFS upper-layer applications need to be started.

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The HDFS backup file save path is correct.
- The HDFS upper-layer applications are stopped.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **HDFS** under **Service Data**.

Step 8 Set **Path Type** of **HDFS** to a backup directory type.

The following backup directory types are supported:

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster.

  If you select **RemoteHDFS**, set the following parameters:

  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

- **NFS**: indicates that backup files are stored in NAS using the NFS protocol. If you select **NFS**, set the following parameters:

  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.

  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **CIFS**: indicates that backup files are stored in NAS using the CIFS protocol. If you select **CIFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select **SFTP**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.
  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.

  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.

Step 9 In the **Backup Data** column of the **Data Configuration** page, select one or more pieces of backup data that needs to be restored based on service requirements. In the **Target Path** column, specify the target location after backup data restoration.

You are advised to set **Target Path** to a new path that is different from the backup path.

Step 10 Click **Verify** to check whether the restoration task is configured correctly.

- If the queue name is incorrect, the verification fails.
- If the specified directory to be restored does not exist, the verification fails.

Step 11 Click **OK**.

Step 12 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

----End

2.3.2.3.6 Recovering Hive Service Data

Scenario

Hive data needs to be recovered in the following scenarios: data is modified or deleted unexpectedly and needs to be restored. After an administrator performs critical data adjustment in the Hive, an exception occurs or the operation has not achieved the expected result. All modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a recovery task in FusionInsight Manager to recover Hive data. Only manual restoration tasks are supported.

Hive backup and restoration cannot identify the service and structure relationships of objects such as Hive tables, indexes, and views. When executing backup and restoration tasks, you need to manage unified restoration points based on service scenarios to ensure proper service running.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To recover data when the service is running properly, you are advised to manually back up the latest management data before recovering data. Otherwise, the Hive data that is generated after the data backup and before the data recovery will be lost.
- To prevent stopping a service or instance from affecting data restoration of other components, do not restore Hive service data and HDFS/HBase metadata at the same time. Otherwise, Hive service data restoration fails.

Impact on the System

- During data restoration, user authentication stops and users cannot create new connections.
- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is recovered, the Hive upper-layer applications need to be started.

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust has been configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Cross-cluster replication has been configured for the active and standby clusters. For details, see **Enabling Cross-Cluster Replication**.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.
- The database for storing restored data tables, the HDFS save path of data tables, and the list of users who can access restored data are planned.
- The Hive backup file save path is correct.
- The Hive upper-layer applications are stopped.
- You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, choose **More** > **View History** to view historical backup task execution records.

In the displayed window, locate a specified success record and click **View** in the **Backup Path** column to view the backup path information of the task and find the following information:

- **Backup Object** specifies the data source of the backup data.
- **Backup Path** specifies the full path where the backup files are saved.

  Select the correct item, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In the **Restoration Configuration** area, select **Hive**.

Step 8 Set **Path Type** of **Hive** to a backup directory type.

The following backup directory types are supported:

- **RemoteHDFS**: indicates that the backup files are stored in the HDFS directory of the standby cluster. If you select **RemoteHDFS**, set the following parameters:
  - **Source NameService Name**: indicates the NameService name of the backup data cluster. You can enter the built-in NameService name of the remote cluster, for example, **haclusterX**, **haclusterX1**, **haclusterX2**, **haclusterX3**, or **haclusterX4**. You can also enter a configured NameService name of the remote cluster.
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Source NameNode IP Address**: indicates the NameNode service plane IP address of the standby cluster, supporting the active node or standby node.
  - **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select a Hive backup file set that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **NFS**: indicates that backup files are stored in NAS using the NFS protocol. If you select **NFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.

  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select a Hive backup file set that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **CIFS**: indicates that backup files are stored in NAS using the CIFS protocol. If you select **CIFS**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the NAS server.
  - **Port**: indicates the port number used to connect to the NAS server over the CIFS protocol. The default value is **445**.
  - **Username**: indicates the username set when the CIFS protocol is configured.
  - **Password**: indicates the password set when the CIFS protocol is configured.
  - **Source Path**: indicates the full path of the backup file on the NAS server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select a Hive backup file set that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **100**.
- **SFTP**: indicates that backup files are stored in the server using the SFTP protocol.

  If you select **SFTP**, set the following parameters:
  - **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
  - **Server IP Address**: indicates the IP address of the server where the backup data is stored.

  - **Port**: indicates the port number used to connect to the backup server over the SFTP protocol. The default value is **22**.
  - **Username**: indicates the username for connecting to the server using the SFTP protocol.
  - **Password**: indicates the password for connecting to the server using the SFTP protocol.
  - **Source Path**: indicates the full path of the backup file on the backup server, for example, *Backup path/Backup task name_Data source_Task creation time*.
  - **Queue Name**: indicates the name of the Yarn queue used for backup task execution.
  - **Recovery Point List**: Click **Refresh** and select an HDFS directory that has been backed up in the standby cluster.
  - **Target NameService Name**: indicates the NameService name of the backup directory. The default value is **hacluster**.
  - **Maximum Number of Maps**: indicates the maximum number of maps in a MapReduce task. The default value is **20**.
  - **Maximum Bandwidth of a Map (MB/s)**: indicates the maximum bandwidth of a map. The default value is **1**.

Step 9 Set **Backup Data** in the **Data Configuration** to one or multiple backup data sources to be recovered based on service requirements. In the **Target Database** and **Target Path** columns, specify the target database and file save path after backup data recovery.

Configuration restrictions:

- Data can be restored to the original database, but data tables must be stored in a new path that is different from the backup path.
- To restore Hive index tables, select the Hive data tables that correspond to the Hive index tables to be restored.
- If a new restoration directory is selected to avoid affecting the current data, HDFS permission must be manually granted so that users who have permission of backup tables can access this directory.
- Data can be restored to other databases. In this case, HDFS permission must be manually granted so that users who have permission of backup tables can access the HDFS directory that corresponds to the database.

Step 10 Set **Force recovery** to **true**, which indicates to forcibly recover all backup data when a data table with the same name already exists. If the data table contains new data added after backup, the new data will be lost after the data recovery. If you set the parameter to **false**, the restoration task is not executed if a data table with the same name exists.

Step 11 Click **Verify** to check whether the restoration task is configured correctly.

- If the queue name is incorrect, the verification fails.
- If the specified directory to be restored does not exist, the verification fails.
- If the forcibly replacement conditions are not met, the verification fails.

Step 12 Click **OK**.

Step 13 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

----End

2.3.2.3.7 Restoring IoTDB Service Data

Scenario

IoTDB service data needs to be restored in the following scenarios: Data is modified or deleted unexpectedly and needs to be restored. After an administrator performs major operations (such as upgrade and data adjustment) on IoTDB, an exception occurs or the expected result is not achieved. All modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create an IoTDB restoration task on FusionInsight Manager. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To restore the data when services are normal, manually back up the latest management data first and then restore the data. Otherwise, the IoTDB data that is generated after the data backup and before the data restoration will be lost.

Impact on the System

- During data restoration, user authentication stops and users cannot create new connections.
- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is restored, the IoTDB upper-layer applications need to be started.

Prerequisites

- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.

- The IoTDB backup file save path is correct.
- The IoTDB upper-layer applications are stopped.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the row where the specified backup task is located, choose **More** > **View History** in the **Operation** column to display the historical execution records of the backup task.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.
- **Backup Path**: indicates the full path where backup files are stored.

  Select the correct path, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In **Restoration Configuration**, choose **IoTDB** > **IoTDB** under **Service Data**.

Step 8 Set **Path Type** of **IoTDB** to a backup directory type.

The following backup directory types are supported:

RemoteHDFS: indicates that the backup files are stored in the HDFS directory of the standby cluster.

If you select this option, configure the following parameters:

- **Source NameService Name**: indicates the NameService name of the backup data cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
- **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
- **Source Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
- **Source Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
- **Source NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the standby cluster.
- **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time*.

- **Recovery Point List**: Click **Refresh** and select an IoTDB directory that has been backed up in the standby cluster.

Step 9 In the **Backup Data** column of the **Data Configuration** page, select one or more pieces of backup data that needs to be restored based on service requirements. In the **Target Path** column, specify the target location after backup data restoration.

You are advised to set **Target Path** to a new path that is different from the backup path.

Step 10 Set **Force recovery** to **true**, which indicates that all backup data is forcibly restored when a data table with the same name already exists. If the data table contains new data added after backup, the new data will be lost after the data restoration. If you set the parameter to **false**, the restoration task is not executed if a data table with the same name exists.

Step 11 Click **Verify** to check whether the restoration task is configured correctly.

- If the queue name is incorrect, the verification fails.
- If the specified directory to be restored does not exist, the verification fails.

Step 12 Click **OK**.

Step 13 In the restoration task list, locate the row where the created task is located, and click **Start** in the **Operation** column.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

----End

2.3.2.3.8 Restoring MOTService Service Data

Scenario

Restore MOTService data in the following scenarios: Data is modified or deleted accidentally, or needs to be recovered; data exceptions occur or the change results are not as expected after major operations such as upgrade or data modification are performed on MOTService; all modules are faulty and become unavailable; data is migrated to a new cluster.

Users can create MOTService restoration tasks on FusionInsight Manager to restore MOTService data. Only manual restoration tasks are supported.

NOTICE

- Data can be restored only when the system version during data backup is the same as the current system version.
- To restore data when services are normal, manually back up the latest management data before restoring data. Otherwise, the MOTService data generated after the data backup and before the data restoration will be lost.

Impact on the System

- During data restoration, user authentication stops and users cannot create new connections.
- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is restored, the upper-layer applications of MOTService need to be restarted.

Prerequisites

- You have prepared a standby cluster if you need to restore data remotely from HDFS. If the active cluster is deployed in security mode (with Kerberos authentication enabled) and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode (with Kerberos authentication disabled), mutual trust is not required.
- The time on the active and standby clusters must be the same, and the NTP service on the active and standby clusters uses the same time source.
- The database for storing restored data tables, the location for storing the data tables in HDFS, and the list of users who can access the restored data have been planned.
- The location for storing MOTService backup files is correct.
- The upper-layer applications of MOTService have been stopped.
- You have logged in to FusionInsight Manager by referring to **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of a specified task in the task list, click **More** and select **View History** to view historical execution records of backup tasks.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.
- **Backup Path**: indicates the full path for storing backup files.

  Select the correct path and copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In **Restoration Configuration**, select **MOTService** under **Service data**.

Step 8 Set **Path Type** of **MOTService** to a backup directory type.

Currently, only the **RemoteHDFS** type is available.

RemoteHDFS: indicates that backup files are stored in HDFS of the standby cluster. If you select this option, you also need to configure the following parameters:

- **Source NameService Name**: indicates the NameService name of the backup data cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
- **IP Mode**: indicates the mode of the target IP address. The system automatically selects an IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
- **Source Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
- **Source Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
- **Source NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the destination cluster.
- **Source Path**: indicates the full path of the HDFS directory for storing backup data of the standby cluster. For details, see **Backup Path** obtained in **Step 2**, for example, *Backup path/Backup task name_Data source_Task creation time/*.

Step 9 Click **OK**.

Step 10 Choose **Cluster** > **Services** > **MOTService**. On the **Dashboard** page that is displayed, click **Stop** to stop the MOTService service as prompted.

Step 11 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task. After the restoration task is complete, manually start the MOTService service.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be re-executed.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to re-execute the task.

----End

2.3.2.3.9 Restoring Solr Service Data

Scenario

Solr service data needs to be restored in the following scenarios: Data is modified or deleted unexpectedly and needs to be restored. After an administrator performs major operations (such as upgrade and data adjustment) on Solr, an exception occurs or the expected result is not achieved. All modules are faulty and become unavailable. Data is migrated to a new cluster.

System administrators can create a Solr service data restoration task on FusionInsight Manager. After the task is successfully executed, the service data is restored. Only manual restoration tasks are supported.

NOTICE

- Data restoration can be performed only when the system version is consistent with that during data backup.
- To restore Solr service data when services are normal, manually back up the latest service data first and then restore the data. Otherwise, the Solr service data that is generated after the data backup and before the data restoration will be lost.
- During the execution of a restoration task, if the index to be restored already exists, the index will be deleted and then automatically created. Therefore, index-related service operations may be affected during the restoration task execution.
- Ensure that the running status of all instances in the cluster is normal and can receive requests properly. To ensure successful restoration, do not perform operations such as adding, deleting, stopping, or restarting Solr instances, stopping or restarting the Solr service, or stopping or restarting the cluster.
- Solr metadata and service data cannot be restored at the same time. Otherwise, service data restoration fails.

Impact on the System

- After the data is restored, the data generated after the data backup and before the data restoration is lost.
- After the data is restored, the Solr upper-layer applications need to be started.

Prerequisites

- The Solr backup file save path is correct.
- The Solr upper-layer applications are stopped.
- If you need to restore data from a remote HDFS, prepare a standby cluster. If the active cluster is deployed in security mode and the active and standby clusters are not managed by the same FusionInsight Manager, mutual trust must be configured. For details, see **Configuring Cross-Manager Mutual Trust Between Clusters**. If the active cluster is deployed in normal mode, no mutual trust is required.
- Time is consistent between the active and standby clusters and the NTP services on the active and standby clusters use the same time source.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the **Operation** column of the specified task in the task list, choose **More** > **View History**.

In the window that is displayed, select a success record and click **View** in the **Backup Path** column to view its backup path information and find the following information:

- **Backup Object**: indicates the backup data source.

- **Backup Path**: indicates the full path where backup files are stored.

  Select the correct path, and manually copy the full path of backup files in **Backup Path**.

Step 3 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Restoration Management**.

Step 4 Click **Create**.

Step 5 Set **Task Name** to the name of the restoration task.

Step 6 Select the desired cluster from **Recovery Object**.

Step 7 In **Restoration Configuration**, choose **Solr** > **Solr** under **Service Data**.

Step 8 Set **Path Type** of **Solr** to a restoration directory type.

The following types of directories can be restored:

RemoteHDFS: indicates that the backup files are stored in the HDFS directory of the standby cluster.

If you select this option, configure the following parameters:

- **Source NameService Name**: indicates the NameService name of the backup data cluster, for example, **hacluster**. You can obtain it from the **NameService Management** page of HDFS of the standby cluster.
- **IP Mode**: indicates the mode of the target IP address. The system automatically selects the IP address mode based on the cluster network type, for example, **IPv4** or **IPv6**.
- **Source Active NameNode IP Address**: indicates the service plane IP address of the active NameNode in the standby cluster.
- **Source Standby NameNode IP Address**: indicates the service plane IP address of the standby NameNode in the standby cluster.
- **Source NameNode RPC Port**: indicates the value of **dfs.namenode.rpc.port** in the HDFS basic configuration of the standby cluster.
- **Source Path**: indicates the full path of HDFS directory for storing backup data of the standby cluster, for example, *Backup path/Backup task name_Data source_Task creation time*.
- **Recovery Point List**: Click **Refresh** and select a Solr snapshot directory that has been backed up in the standby cluster.

Step 9 In the **Data Configuration** area, select one or more indexes for **Select Data** based on service requirements.

Step 10 Specify **Force recovery**. The value **false** does not take effect. All backup data is forcibly restored when there are indexes with the same name. If the index contains data added after backup, the added data will be lost after the data restoration.

Step 11 Click **Verify** to check whether the restoration task is configured correctly.

- If the specified directory to be restored does not exist, the verification fails.
- If the forcible restoration conditions are not met, the verification fails.

Step 12 Click **OK**.

Step 13 In the restoration task list, locate the row where the created task is located, and click **Start** in the **Operation** column.

- After the restoration is successful, the progress bar is in green.
- After the restoration is successful, the restoration task cannot be executed again.
- If the restoration task fails during the first execution, rectify the fault and click **Retry** to execute the task again.

----End

2.3.2.4 Restoring OMS Data from a Third-Party Server Outside a Cluster

Scenario

Restore OMS data from a third-party server.

NOTE

This section applies only to physical machine clusters.

Impact on the System

Some services of the cluster are unavailable during the restoration.

Procedure

Step 1 Copy the OMS backup file (for example, **X.X.X_OMS_20170422110055.tar.gz**) from the third-party server to the directory on the active management node.

Step 2 Restore Manager (including OMS and LDAP) and DBService data in LocalDir mode by referring to **Restoring OMS Data** and **Recovering DBService Data**.

----End

2.4 Common Operation

2.4.1 Managing Local Quick Recovery Tasks

Scenario

When DistCp is used to back up data, the backup snapshot is saved to HDFS of the active cluster. FusionInsight Manager supports using the local snapshot for quick data restoration, requiring less time than restoring data from the standby cluster.

Use FusionInsight Manager and the snapshots on the HDFS of the active cluster to create a local quick restoration task and execute the task.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the backup task list, locate a created task and click **Restore** in the **Operation** column.

Step 3 Check whether the system displays "No data is available for quick restoration. Create a task on the restoration management page to restore data".

- If yes, click **OK** to close the dialog box. No backup data snapshot is created in the active cluster, and no further action is required.
- If no, go to **Step 4** to create a local quick restoration task.

NOTE

- Only Hive, HBase, and HDFS support quick restoration.
- Metadata cannot be quickly restored.

Step 4 Set **Name** to the name of the local quick restoration task.

Step 5 Select a data source in the **Restoration Configuration** area.

Step 6 Set **Recovery Point List** to a recovery point that contains the backup data.

Step 7 Set **Queue Name** to the name of the Yarn queue used in the task execution. The name must be the same as the name of the queue that is running properly in the cluster.

Step 8 Set **Data Configuration** to the object to be recovered.

Step 9 Click **Verify**, and wait for the system to display "The restoration task configuration is verified successfully."

Step 10 Click **OK**.

Step 11 In the restoration task list, locate a created task and click **Start** in the **Operation** column to execute the restoration task.

After the task is complete, **Task Status** of the task is displayed as **Successful**.

----End

2.4.2 Modifying a Backup Task

Scenario

This section describes how to modify the parameters of a created backup task on FusionInsight Manager to meet changing service requirements. The parameters of restoration tasks can only be viewed but cannot be modified.

Impact on the System

After a backup task is modified, the new parameters take effect when the task is executed next time.

Prerequisites

- A backup task has been created.
- A new backup task policy has been planned based on the actual situation.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**.

Step 2 In the task list, locate a specified task, click **Configure** in the **Operation** column to go to the configuration modification page.

On the displayed page, modify the following parameters:

- Started
- Period
- Destination NameService Name
- Target NameNode IP Address
- Target Path
- Max Number of Backup Copies
- Maximum Number of Recovery Points
- Maximum Number of Maps
- Maximum Bandwidth of a Map

NOTE

After the **Target Path** parameter of a backup task is modified, this task will be performed as a full backup task for the first time by default.

Step 3 Click **OK** to save the settings.

----End

2.4.3 Viewing Backup and Recovery Tasks

Scenario

This section describes how to view created backup and recovery tasks and check their running status on FusionInsight Manager.

Prerequisites

You have logged in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Procedure

Step 1 On FusionInsight Manager, choose **O&M** > **Backup and Restoration**.

Step 2 Click **Backup Management** or **Restoration Management**.

Step 3 In the task list, obtain the previous execution result in the **Task Status** and **Task Progress** column. Green indicates that the task is executed successfully, and red indicates that the execution fails.

Step 4 In the **Operation** column of a specified task in the task list, choose **More** > **View History** or click **View History** to view the historical record of backup and restoration task execution.

In the displayed window, click ⌄ before a specified record to display log information about the execution.

----End

Related Tasks

- Starting a backup or restoration task

  In the task list, locate a specified task and choose **More** > **Back Up Now** or click **Start** in the **Operation** column to start a backup or restoration task that is ready or fails to be executed. Executed restoration tasks cannot be repeatedly executed.

- Stopping a backup or restoration task

  In the task list, locate a specified task and choose **More** > **Stop** or click **Stop** in the **Operation** column to stop a backup or restoration task that is running. After the task is successfully stopped, its **Task Status** changes to **Stopped**.

- Deleting a backup or restoration task

  In the task list, locate a specified task and choose **More** > **Delete** or click **Delete** in the **Operation** column to delete a backup or restoration task. Backup data will be reserved by default after a task is deleted.

- Suspending a backup task

  In the task list, locate a specified task and choose **More** > **Suspend** in the **Operation** column to suspend a backup task. Only periodic backup tasks can be suspended. Suspended backup tasks are no longer executed automatically. When you suspend a backup task that is being executed, the task execution stops. To resume a task, choose **More** > **Resume**.

2.4.4 Creating a Mirror Cluster Using Backup Data

Scenario

You can create a mirror cluster of the active cluster using backup data when all nodes in the active cluster are offline due to exceptions and cannot provide services, or a test cluster that is completely the same as the active cluster needs to be created.

This section describes how to create a test cluster that is the same as the primary cluster in a new network environment.

NOTE

This section applies only to physical machine clusters.

Prerequisites

- The network planning of the mirror cluster is the same as that of the active cluster.
- The number of hosts in the mirror cluster is the same as that in the active cluster. The host names are the same as those in the active cluster.
- The host disks and partitions in the mirror cluster are the same as those in the active cluster.

- All IP addresses, including the management, service, and floating IP addresses of the nodes in the mirror cluster, are the same as the corresponding IP addresses of the active cluster.
- All nodes in the active cluster are offline and cannot provide services if the active cluster and the mirror cluster are connected over a network. Or the mirror cluster is to be installed in an isolated network environment.
- The OMS, DBService, Kafka, and NameNode metadata in the active cluster has been backed up to the standby cluster using DistCp, the HBase, Hive, and HDFS service data has been backed up to the standby cluster using DistCp, and the data can be accessed properly. The backup data has been uploaded to and saved in the nodes in the mirror cluster.

Procedure

Step 1 Select nodes to install the OS of the same version of that on nodes in the active cluster, and run the **preinstall** script to perform OS preconfiguration and partition mounting.

For details, see *MRS Installation Guide in Huawei Cloud Stack 8.3.1 Software Installation Guide for gPaaS & AI DaaS Services* . All host names, IP addresses, the number of mounted disks, and disk space must be the same as those of the active cluster.

Step 2 Install Manager on the management nodes in the mirror cluster.

On the nodes where the **preinstall** script is executed, install the active and standby Managers by referring to "Installing Manager on Two Management Nodes" in *MRS Installation Guide in Huawei Cloud Stack 8.3.1 Software Installation Guide for gPaaS & AI DaaS Services* . The installation user must be the same as that of the active cluster, such as user **root** or **omm**.

NOTE

- The key values in the installation scripts must be the same as those used by the original cluster.
- If the HA root certificate has been replaced, you need to perform the following operations to add a certificate after the software package is decompressed to the node where the certificate has been replaced (for example, the software package is decompressed to **/opt**).
  1. Go to the **/opt/FusionInsight_Manager/software/hasslCert** directory.

     cd /opt/FusionInsight_Manager/software/hasslCert
  2. Prepare a root certificate and place the **root-ca.crt** HA root certificate file and the **root-ca.pem** key file in the **/opt/FusionInsight_Manager/software/hasslCert** directory.

     The certificate must be the same as the original one.
  3. Set **password** (empty by default) in the **/opt/FusionInsight_Manager/software/hasslCert/hasslCert.ini** file to the password of the root certificate generated. Configuration files containing authentication passwords pose security risks. Delete such files after configuration or store them securely.

     For example, set **password** to *Password*. After the cluster is installed, the password will be deleted.

Step 3 Log in to the two management nodes as user **omm**, and run the following command to update the **/etc/hosts** file on the nodes:

sh ${BIGDATA_HOME}/om-server/om/sbin/updateOMSHosts.sh

The command is run successfully if the following information is displayed:

```
Succeed to update /etc/hosts.
```

Step 4 Log in to the active OMS node as user **omm**, and run the following commands to use the data backed up before the fault occurs to restore the OMS data. The domain of the OMS data before the fault is backed up needs to be the same as that of the newly installed environment. You can view the status of the node by running the **$CONTROLLER_HOME/sbin/status-oms.sh** script after logging in to the OMS node.

cd ${CONTROLLER_HOME}/sbin

sh huaweibigdata_backup.sh -r *Full path of OMS backup files*

Step 5 Run the following commands on the active OMS node to disable the active/standby OMS switchover:

cd ${OMS_RUN_PATH}/workspace/ha/module/hacom/tools/

./ha_client_tool --forbidswitch --name=product --time=120

Step 6 Log in to the active management node as user **omm** and run the following command to check the creation time of a common **default-oms** backup task in a cluster:

gsql -p *port* **-W** *password* **-U omm -d omm -c "select * from TBL_BACKUP_TASKINFO where TASKNAME='default-oms'"**

Example:

```
TASKNAME   | CREATETIME   | TASKTYPE |  FIRSTBACKUP  | PERIODTIME | COUNTER | STRATEGY |
CLUSTERID | LOCKED
-------------+---------------+----------+---------------+------------+---------+----------+-----------+--------
default-oms | 1567018790868 | PERIOD   | 1567094400868 |       3600 |      0  |       1 |         0 | f
(1 row)
```

Record the value of **CREATETIME** in the command output.

Step 7 Log in to the active and standby OMS nodes as user **omm**, and run the following commands to restart OMS:

cd ${CONTROLLER_HOME}/sbin

./restart-oms.sh

Run the **sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh** command to check whether the value of **HAAllResOK** of the active management node is **Normal**. If yes, OMS is restarted successfully.

Step 8 Wait for several minutes, log in to FusionInsight Manager, choose **O&M** > **Backup and Restoration** > **Backup Management**, and check whether the backup management page is displayed properly.

- If yes, go to **Step 11**.
- If no, go to **Step 9**.

Step 9 Log in to the active management node as user **omm** and run the following command to restore the backup management page. To obtain the default password of the OMS database user **omm**, see ***Huawei Cloud Stack 8.3.1***

Account List or contact the system administrator. The default port number is **20015**.

gsql -p *port* **-W** *password* **-U omm -d omm -c "select * from TBL_BACKUP_TASKINFO where TASKNAME='default-oms'"**

Example:

```
TASKNAME   | CREATETIME   | TASKTYPE |  FIRSTBACKUP  | PERIODTIME | COUNTER | STRATEGY |
CLUSTERID | LOCKED
-------------+---------------+----------+---------------+------------+---------+----------+------------+--------
default-oms | 1567018790868 | PERIOD   | 1567094400868 |       3600 |      0  |       1 |         0 | f
default-oms | 1566018790869 | PERIOD   | 1567094400868 |       3600 |     45 |       1 |         0 | f
(2 rows)
```

Delete the redundant **default-oms** tasks whose **create_time** is **1567018790868** (time queried in **Step 6**). The data backed up before the fault occurs indicates the **default-oms** task time is **1566018790869**.

gsql -p *port* **-W** *password* **-U omm -d omm -c "delete from TBL_BACKUP_TASKINFO where TASKNAME='default-oms' and CREATETIME=***create_time***"**

gsql -p *port* **-W** *password* **-U omm -d omm -c "delete from TBL_BACKUP_RECORDS where TASKNAME='default-oms' and CREATETIME=***create_time***"**

gsql -p *port* **-W** *password* **-U omm -d omm -c "delete from TBL_BACKUP_CONFIGS where TASKNAME='default-oms' and CREATETIME=***create_time***"**

Step 10 Log in to the active management node as user **omm** and run the following command to restart the Controller service:

sh $CONTROLLER_HOME/sbin/restart-controller.sh

Step 11 Log in to the active management node as user **omm** and run the following commands to replace the NodeAgent key files of the active and standby management nodes. The NodeAgent on the active and standby nodes is not reinstalled. Therefore, no corresponding certificate files are generated.

cp -rp $CONTROLLER_HOME/security/cert/subcert/certFile ${NODE_AGENT_HOME}/security/cert/subcert/

cp -p $CONTROLLER_HOME/security/cert/subcert/certFile/password.property ${NODE_AGENT_HOME}/security/cert/subcert/certFile/password.file

cp -p $CONTROLLER_HOME/etc/om/security.properties ${NODE_AGENT_HOME}/etc/agent/

cp -p $CONTROLLER_HOME/packaged-distributables/tomcat.crt ~/

scp -rp $CONTROLLER_HOME/security/cert/subcert/certFile omm@*IP address of the standby node***:${NODE_AGENT_HOME}/security/cert/subcert/**

scp -p $CONTROLLER_HOME/security/cert/subcert/certFile/password.property omm@*IP address of the standby node***:${NODE_AGENT_HOME}/security/cert/subcert/certFile/password.file**

scp -p $CONTROLLER_HOME/etc/om/security.properties omm@*IP address of the standby node***:${NODE_AGENT_HOME}/etc/agent/**

scp -p $CONTROLLER_HOME/packaged-distributables/tomcat.crt omm@*IP address of the standby node***:~/**

Step 12 Reinstall all hosts in the cluster by referring to **Reinstalling a Host**.

After the reinstallation, the hosts, topologies, configuration, services, and instances of the mirror cluster are the same as those of the active cluster.

Step 13 After the cluster host is reinstalled, use the files backed up before the fault occurs to restore the LdapServer data in the cluster by referring to **Restoring OMS Data**.

Step 14 Log in to the active OMS node as user **omm**, and run the following commands to enable the active/standby OMS switchover:

cd ${OMS_RUN_PATH}/workspace/ha/module/hacom/tools/

./ha_client_tool --cancelforbidswitch --name=product

Step 15 Restore DBService data.

Restore the DBService data in the cluster by referring to **Recovering DBService Data**.

Step 16 Restore the HBase, HDFS, and Hive data.

----End

2.5 Appendix

2.5.1 Solution to the Situation Where MRS-MySQL Is Not Registered with CloudCMDB

Scenarios

This solution is only used to check whether MRS-MySQL is registered with CloudCMDB after MRS 3.0.2-LTS.2 or an earlier version is upgraded to the current version. If MRS-MySQL is not registered with CloudCMDB, you need to back up MRS database data on ManageOne.

Procedure

Step 1 Obtain the IP address of the MRS-DB node.

On Service OM, choose **Services** > **Resource** > **Compute Resource**, search for the **MRS-DB** node on the **VMs** tab page, and record its IP address.

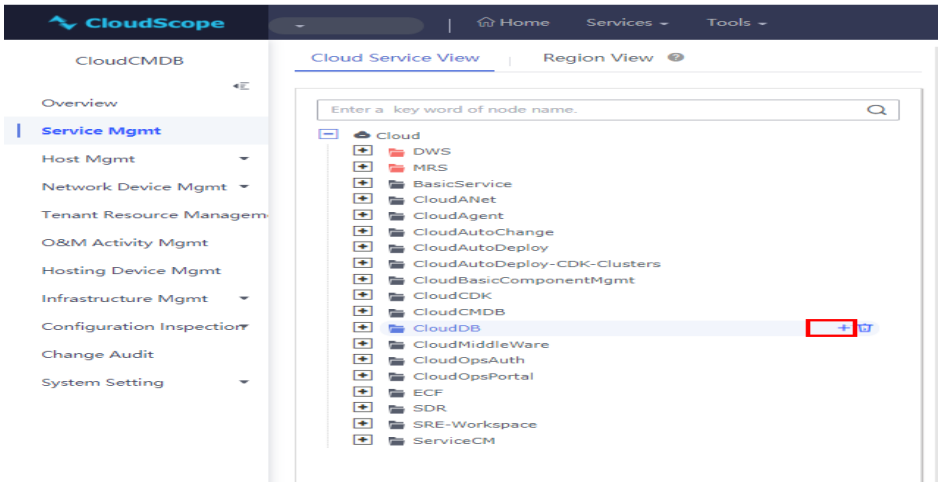
Step 2 Log in to CloudScope as user **op_cdk_sso** and choose **Services** > **Change Mgmt** > **CloudCMDB**.

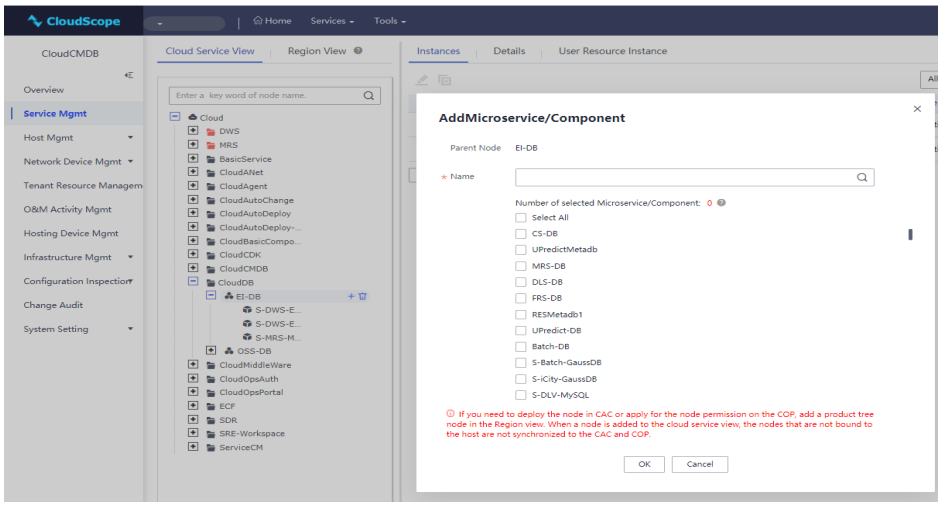
Step 3 On the CloudCMDB page, click **Service Mgmt**. In the **Cloud Service View** area, search for **CloudDB** and check whether **S-MRS-MySQL** exists under **CloudDB > EI-DB**.

- If yes, MRS-MySQL has been registered with CloudCMDB. No further action is required.

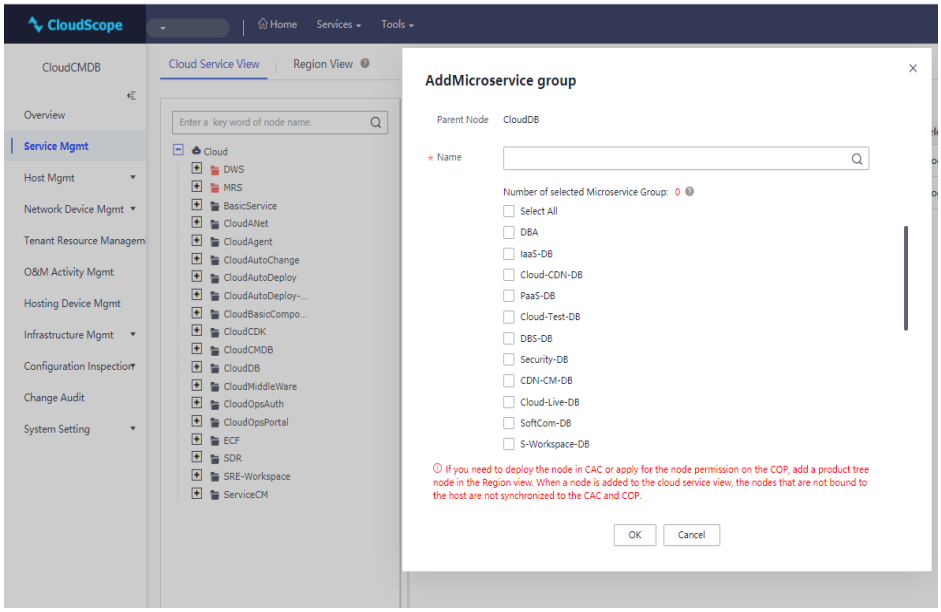
- If no, go to **Step 4**.

Step 4 Click the plus sign (+) on the right of **CloudDB**. In the displayed **Add Microservice Group** dialog box, search for and select **EI-DB**, and click **OK**.

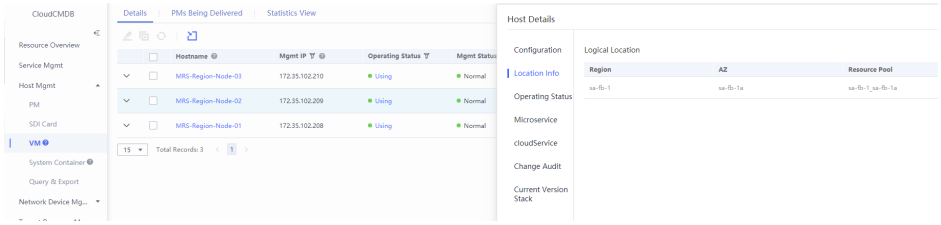




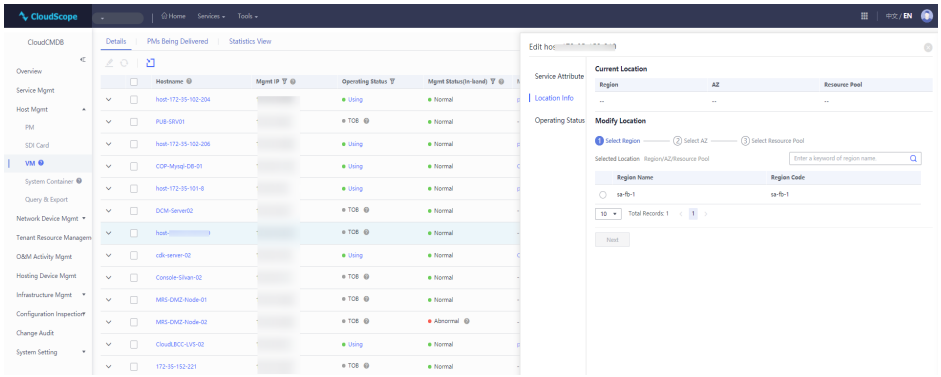
Step 5 Click the plus sign (+) on the right of **EI-DB**. In the displayed dialog box, search for and select **S-MRS-MySQL**, and click **OK**.



Step 6 On the CloudCMDB page, choose **Host Mgmt** > **VM**, select **Host name** from the search box, and search for **MRS-Region-Node**. Select the host whose **Operating Status** is **Using** and **Mgmt Status** is **Normal**, and click the host name. On the **Host Details** page, click **Location Info** to view and save the location information.

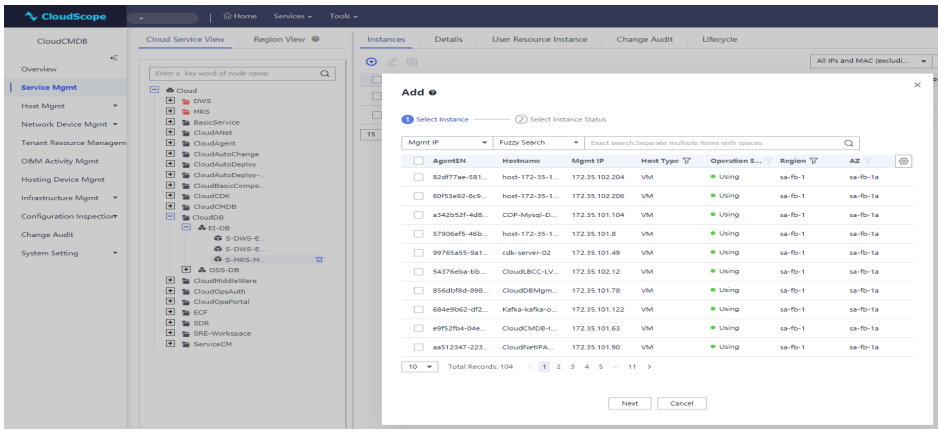


Step 7 Choose **Host Mgmt** > **VM** and search for "management IP (accurate)" in the search box based on the IP address of the MRS-DB node obtained in **Step 1**. In the **Operation** column of the VM, click  . In the displayed dialog box, click **Location Info** and perform operations as prompted and based on the location information in **Step 6**.



Step 8 On the CloudCMDB page, choose **Service Mgmt**. In the **Cloud Service View** area, search for **CloudDB** and choose **CloudDB** > **EI-DB** > **S-MRS-MySQL**. On the **Instances** tab page, click  .

Search for and add the MRS-DB database node based on the IP address of MRS-DB obtained in **Step 1**.



After the VM is added, **Operating Status** of the VM is **Using**, and **Mgmt Status** is **Normal**.

Step 9 Log in to CloudScope as a user with the **CloudAutoChange_manager** role.

Step 10 Choose **Services** > **Change Mgmt** > **CloudAutoChange**, and then choose **Operation Platform** > **Change Instance Schema**.

Step 11 Select **Search by Product Tree**, enter **CloudDB** in the search box, and choose **EI-DB > S-MRS-MySQL**. Search for database nodes based on the IP address of MRS-DB obtained in **Step 1**, and select all the found database nodes.

Set **Batch Policy** to **Auto Batch** and **Batch Setting** to **Continue All**. Select **Customize** for **Change Instance Schema** to and ensure that the schema information is the same as that of MRS-API. Click **OK**.

----End

2.5.2 Enabling Cross-Cluster Replication

Scenario

DistCp is used to replicate the data stored in HDFS from a cluster to another cluster. DistCp depends on the cross-cluster replication function, which is disabled by default. You need to enable it for both clusters.

This section describes how to modify parameters on FusionInsight Manager to enable the cross-cluster replication function. After this function is enabled, you can create a backup task for backing up data to the remote HDFS (RemoteHDFS).

Impact on the System

Yarn needs to be restarted to enable the cross-cluster replication function and cannot be accessed during restart.

Prerequisites

- The **hadoop.rpc.protection** parameter of HDFS in the two clusters for data replication must use the same data transmission mode. The default value is

privacy, indicating encrypted transmission. The value **authentication** indicates that transmission is not encrypted.

- For clusters in security mode, you need to configure mutual trust between clusters.

Procedure

Step 1 Log in to FusionInsight Manager of one of the two clusters.

Step 2 Choose **Cluster** > **Services** > **Yarn** and click **Configurations** then **All Configurations**.

Step 3 In the navigation pane, choose **Yarn(Service)** > **Distcp**.

Step 4 Modify **dfs.namenode.rpc-address**, set **haclusterX.remotenn1** to the service IP address and RPC port of one NameNode instance of the peer cluster, and set **haclusterX.remotenn2** to the service IP address and RPC port number of the other NameNode instance of the peer cluster.

haclusterX.remotenn1 and **haclusterX.remotenn2** do not distinguish active and standby NameNodes. The default NameNode RPC port is 25000 and cannot be modified on Manager.

Examples of modified parameter values: **10.1.1.1:25000** and **10.1.1.2:25000**.

NOTE

- If Federation is configured in the peer cluster with multiple pairs of NameNodes (multiple NameServices), here you can only configure the RPC addresses of the two NameNodes in one of the NameService for the same set of parameters. Configuring the RPC addresses of two NameNodes that do not belong to the same NameService is prohibited.

  ECS/BMS clusters do not support Federation.
- If data of the current cluster needs to be backed up to the HDFS of multiple clusters, you can configure the corresponding NameNode RPC addresses to haclusterX1, haclusterX2, haclusterX3, and haclusterX4.

Step 5 Click **Save**. In the confirmation dialog box, click **OK**.

Step 6 Restart the Yarn service.

Step 7 Log in to FusionInsight Manager of the other cluster and repeat **Step 2** to **Step 6**.

----End

2.5.3 Configuring Cross-Manager Mutual Trust Between Clusters

Scenarios

When two clusters in different security modes need to access each other's resources, the administrator can set up a mutual trust system so that users of external systems can use the system.

The usage range of users in each system is called a domain. Each Manager system must have a unique domain name. Cross-Manager access means users to be used across domains.

Impact on the System

- If the current cluster is a managed physical machine cluster, you need to cancel the management, and then configure cross-cluster mutual trust. Then you can manage the cluster again.
- After cross-cluster mutual trust is configured, users of an external system can be used in the local system. The system administrator needs to periodically check the user rights in the Manager system based on enterprise service and security requirements.
- When configuring cross-cluster mutual trust, you need to stop all clusters, which interrupts services.
- After cross-cluster mutual trust is configured, each of the clusters trusting each other can add Kerberos internal users **krbtgt/***local cluster domain name*@*external cluster domain name* and **krbtgt/***external cluster domain name*@*local cluster domain name*. The two users cannot be deleted. For details about the default password, see ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator. Based on enterprise service and security requirements, the system administrator needs to change the password periodically. The passwords of the four users in the two systems trusting each other must be the same. Connections of cross-Manager service applications may be affected during the password change.
- If the system domain name is changed and there is any running HetuEngine compute instance, restart the compute instance.
- After configuring the cross-cluster mutual trust relationship, download and install the client again for each cluster.

Prerequisites

- The system administrator has specified service requirements and planned domain names of the systems. A domain name can contain uppercase letters, numbers, dots (.), and underscores (_), and must start with a letter or number. For example, **DOMAINA.HW** and **DOMAINB.HW**.
- Before configuring cross-cluster mutual trust, ensure that the domain names of the two Manager systems are different.
- Before cross-cluster mutual trust is configured, ensure that the two systems do not have the same host name or the same IP address.
- Time of two systems configured trust relationships must be consistent and the Network Time Protocol (NTP) service in the two systems must use the same time source.
- Running status of all services in the Manager clusters is **Normal**.
- The **acl.compare.shortName** parameter of the ZooKeeper service of all clusters in the Manager must be set to the default value **true**. Otherwise, change the value to **true** and restart the ZooKeeper service.

Procedure

Step 1 Log in to one FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **Domain and Mutual Trust**.

Step 3 Modify **Peer Mutual Trust Domain**.

Table 2-30 Related parameters

Parameter	Description
realm_name	Enter the domain name of the peer system.
ip_port	Enter the KDC address of the peer system. Value format: *IP address of the node accommodating the Kerberos service in the peer system:Port number* ● In dual-plane networking, enter the service plane IP address. ● If an IPv6 address is used, the IP address must be enclosed in square brackets ([]). ● Use commas (,) to separate the KDC addresses if the active and standby Kerberos services are deployed or multiple clusters in the peer system need to establish mutual trust with the local system. ● You can obtain the port number from the **kdc_ports** parameter of the KrbServer service. The default value is **21732**. To obtain the IP address of the node where the service is deployed, click the **Instance** tab on the KrbServer page and view **Service IP Address** of the KerberosServer role. For example, if the Kerberos service is deployed on nodes at **10.0.0.1** and **10.0.0.2** that have established mutual trust with the local system, the parameter value is **10.0.0.1:21732,10.0.0.2:21732**.

NOTE

If you need to configure mutual trust for multiple Managers, click + to add a new item and set parameters. To delete unnecessary configurations, click −.

Step 4 Click **OK**.

Step 5 Log in to the active management node as user **omm**, and run the following command to update the domain configuration:

sh ${BIGDATA_HOME}/om-server/om/sbin/restart-RealmConfig.sh

The command is successfully executed if the following information is displayed:

```
Modify realm successfully. Use the new password to log in to FusionInsight again.
```

After the restart, some hosts and services cannot be accessed and an alarm is generated. This problem can be automatically resolved in about 1 minute after **restart-RealmConfig.sh** is run.

Step 6 Log in to FusionInsight Manager and restart the cluster or configure expired instances:

Check whether the system domain name of Manager is changed.

- If the system domain name is changed, choose **More** > **Restart** in the upper right corner of the home page, enter the password, select the checkbox for confirming the impact, and click **OK**. Wait until the cluster is restarted.
- If the system domain name is not changed, choose **More** > **Restart Configuration-Expired Instances** in the upper right corner of the home page, enter the password, select the checkbox for confirming the impact, and click **OK**. Wait until the service is restarted.

Step 7 Log out of FusionInsight Manager and then log in again. If the login is successful, the configuration is successful.

Step 8 Log in to the active management node in an ECS/BMS or a managed physical machine cluster as user **omm** and run the following command to update the configurations of the job submission client:

sh /opt/executor/bin/refresh-client-config.sh

Step 9 If a HetuEngine compute instance is running, restart the compute instance.

1. Log in to FusionInsight Manager as a user who accesses the HetuEngine web UI.
2. Choose **Cluster** > **Services** > **HetuEngine**.
3. In the **Basic Information** area on the **Dashboard** page, click the link next to **HSConsole WebUI**. The HSConsole page is displayed.
4. On the **Compute Instance** tab page, click **Stop** in the **Operation** column of a running compute instance. After all compute instances are in the **Stopped** state, click **Start** to restart the compute instance.

Step 10 Log in to the other FusionInsight Manager and repeat the preceding operations.

NOTE

If the current cluster is a managed physical machine cluster, you need to cancel the management, and then configure cross-cluster mutual trust. Then you can manage the cluster again.

----End

2.5.4 How Do I Run Commands or Access Files on Multiple Nodes in a Cluster?

NOTE

This section applies only to physical machine clusters.

Question

During cluster installation or routine maintenance, you can use the script tool in the software package to run a command or access a file on multiple nodes in a cluster.

Answer

Prerequisites

- You have obtained the operator username and password of each node in the cluster and have enabled the user's remote login permission.
- You have decompressed the **FusionInsight_SetupTool_XXX.tar.gz** package, the script tool package used for FusionInsight software installation, to the **/opt** directory of the active management node.

Procedure

Step 1 Log in to the active management node as an operator.

Step 2 Go to the **/opt/FusionInsight_SetupTool/preinstall/tools/cluster** directory and edit the **cluster.ini** file as required.

Table 2-31 Parameters in **cluster.ini**

Parameter	Example Value	Description
g_hosts (mandatory)	192.168.10.[10-20]	Specifies the IP addresses of all nodes where the operation is performed. ● Use commas (,) to separate IP addresses. For example, **192.168.10.10**, **192.168.10.11**. ● Use hyphens (-) to indicate an IP address segment if the IP addresses are consecutive. For example, **192.168.10.[10-20]**. ● Use hyphens (-) to indicate an IP address segment if the IP addresses are consecutive, and use commas (,) to separate IP address segments. For example, **192.168.10.[10-20,30-40]**.
g_user_name	root	Specifies the user who performs the operation.
g_password	N/A	Specifies the password file corresponding to the user who performs the operation. The password is empty by default.

Parameter	Example Value	Description
g_port	22	Specifies the SSH connection port. The default value is **22**.
g_timeout	10	Specifies the SSH connection timeout period. The default value is **10 seconds**. The value increases when the network conditions are poor.
g_ip_model	IPv4	IP address mode of the network where the cluster is located.

By default, **g_password** is left blank. If the passwords of all users who perform the operation are the same, you only need to enter the password once after the command is executed. Otherwise, you must manually create a password file and set **g_password** to the full path of the password file. The password file is in the following format:

```
IP address 1
Password of IP address 1
IP address 2
Password of IP address 2
...
```

NOTE

- Node passwords stored locally have security risks. Therefore, it is recommended that the password of the executor for each node be the same. If you have to use these passwords, ensure that only executors have read and write permissions for the password file and delete the password file immediately after use.
- The password file must be in the UNIX format.
- The last line of the password file cannot be empty.
- Special characters contained in the password do not need to be converted.

For example, the content for creating the **secret.txt** file is as follows:

```
10.10.37.[10-11]
123456!654321
10.10.37.12
123456!
```

Step 3 Run the command based on the scenarios.

1. This command is executed on each node.

   Command format: **./clustercmd.sh** *Detailed command*

   Example (Running the **hostname** command on each node):

   ```
   dc-rack1007-4m:/cluster # ./clustercmd.sh hostname
   ==>>10.10.37.10
   dc-rack1007-1
   ==>>10.10.37.11
   dc-rack1007-2
   ```



```
==>>10.10.37.12==
dc-rack1007-3
```

2. Copy the file from each node to the specified directory on the node.

   Command format: **./clusterscp.sh get** *Source path Target path*

   Example (Copying the **/opt/test/mem.txt** file from each node to the **/opt/result** directory on the node):

```
dc-rack1007-4m:/cluster # ./clusterscp.sh get /opt/test/mem.txt /opt/result
get /opt/result/10.10.37.10_mem.txt from 10.10.37.10:/opt/test/mem.txt successfully.
get /opt/result/10.10.37.11_mem.txt from 10.10.37.11:/opt/test/mem.txt successfully.
get /opt/result/10.10.37.12_mem.txt from 10.10.37.12:/opt/test/mem.txt successfully.
```

3. Copy the specified files or folders from the node to the specified directory on each node.

   Command format: **./clusterscp.sh put** *Source path Target path*

   Example (Copying the **/opt/test/hosts** file from the node to the **/etc** directory on each node):

```
dc-rack1007-4m:/opt/cluster # ./clusterscp.sh put /opt/test/hosts /etc
put /opt/test/hosts to 10.10.37.10:/etc successfully.
put /opt/test/hosts to 10.10.37.11:/etc successfully.
put /opt/test/hosts to 10.10.37.12:/etc successfully.
```

----End

2.5.5 Logging In to FusionInsight Manager

Scenario

Log in to FusionInsight Manager using an account.

Procedure

Step 1 Obtain the URL for logging in to FusionInsight Manager.

Step 2 On login page, enter the username and password.

The default username is **admin**. To obtain the default password, see ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator.

Step 3 Change the password upon your first login.

The password must meet the following complexity requirements:

- Contains 8 to 64 characters.
- Contains at least four types of the following: uppercase letters, lowercase letters, numbers, spaces, and special characters (`~!@#$%^&*()-_=+|[{}];',<.>/\?).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be the same as the current password.

Step 4 Move the cursor over  in the upper right corner of home page, and choose **Logout** from the drop-down list. In the dialog box that is displayed, click **OK** to log out of the current user.

----End

2.5.6 Reinstalling a Host

Scenario

When handling a host fault, you can reinstall all service roles on the host. After the service roles are reinstalled, the service roles on the host are consistent with those displayed in the service topology during cluster installation. If the host to be reinstalled is an isolated host, the host remains in the isolated state after the reinstallation. After the host is reinstalled, service role instance statuses remain unchanged.

This section applies only to physical machine clusters.

Prerequisites

- Before reinstallation, permissions have been assigned to user **omm** if this user is used to reinstall hosts.
- The SSH ports of the nodes in the cluster must be the same. Otherwise, the operations in this section will fail.
- Before reinstallation, check whether the **${BIGDATA_HOME}/om-server/om/etc/om/known_hosts** file on the active and standby OMS nodes contains the IP address of the host you want to reinstall.

  If the IP address is contained, delete the IP address from the **known_hosts** file on the active and standby OMS nodes, and run the **sh ${BIGDATA_HOME}/om-server/om/sbin/restart-controller.sh** command on the active node to restart the controller process.

  Restarting this process does not interrupt services. However, Manager REST interfaces will be unavailable for a short period and will be restored in approximately 1 minute.

NOTE

- If the host to be reinstalled contains the Elasticsearch service, only one EsMaster instance can run on the host. If there are multiple EsMaster instances, the Elasticsearch service will fail.
- If the hosts to be reinstalled contain the IoTDB service, the hosts can only be reinstalled one by one. During the reinstallation, do not restart the cluster, hosts, services, or instances. Otherwise, the IoTDB service will fail.
- If the default user of ClickHouse in common mode **default** and **clickhouse** change the default passwords, the passwords of **default** and **clickhouse** users of the reinstalled ClickHouseServer node will also be reset. In this case, you need to change the password.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Click **Hosts**.

Step 3 Select the check box of the host to be reinstalled.

Step 4 Select **Reinstall** from the **More** drop-down list.

In the displayed dialog box, enter the password of the current login user and click **OK**.

Step 5 In the confirmation dialog box that is displayed, select **I want to delete the selected hosts and accept the consequences of possible service failures and data loss**.

If you do not need to retain the service data on the host, select **Clear Data**.

Step 6 Select user **root**, enter the password, and click **OK**.

After the system displays "Operation succeeded", click **Finish**.

NOTE

- If you reinstall a host as user **omm**, delete the operation rights from user **omm** after the task is complete.
- If the IoTDB service is installed on the reinstalled host, perform the following operations after the host is reinstalled: choose **Cluster** > **Services**, locate the row that contains the IoTDB service, and choose **More** > **Synchronize Configuration** to synchronize the IoTDB configuration.
- In a managed physical machine cluster, if the host of a non-OMS management node is reinstalled, and JobServer instances of the JobGateway service are deployed on the node before the host reinstallation, run the following script after the reinstallation as user **omm** on the standby OMS node to update the client:

  sh ${BIGDATA_HOME}/om-server/om/inst/cloud/linkHCS/repairHCS.sh *IP address of the host to be reinstalled*

----End

2.5.7 How Do I Delete the IDs of Users omm and ommdba from the /etc/uid_list File?

Question

If the OS of the current cluster is Kylin V10, how do I clear the IDs of users **omm** and **ommdba** in the **/etc/uid_list** file before adding and reinstalling a host, or replacing a node?

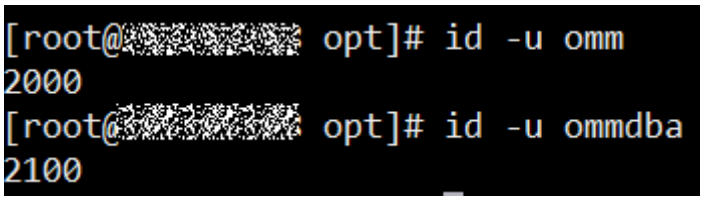
Answer

If the OS of the current cluster is Kylin V10, you can perform the following operations to clear the IDs of users **omm** and **ommdba** before adding and reinstalling a host, or replacing a node.

1. Log in to each node in the cluster as user **root** and run the following commands to query the IDs of users **omm** and **ommdba**:

   id -u omm

   id -u ommdba

   As shown in the preceding figure, the IDs of users **omm** and **ommdba** are **2000** and **2100**, respectively.

2. View and delete the IDs of users **omm** and **ommdba** from the **/etc/uid_list** file. For example, the values of **uid** are **2000** and **2100**.

   vi /etc/uid_list

   If the **/etc/uid_list** file does not exist or the value of uid does not contain **2000** or **2100**, skip this step.

   ```
   [root@xx-xx-xx-xx]# vi /etc/uid_list
   0
   1
   ......
   38
   2000
   2100
   ```

2.5.8 How Do I Prepare for Restoring RemoteHDFS Tasks After Elasticsearch Is Reinstalled?

If the current restoration type is **RemoteHDFS** and Elasticsearch is deleted and then added again when you need to perform the operation described in **Restoring Elasticsearch Service Data**, manually create a snapshot by referring to this section and then perform the restoration task.

Step 1 Log in to the node where the Elasticsearch client is located as user **root**.

Step 2 Run the following commands to go to the client installation directory and set the environment variable:

cd *Client installation path*

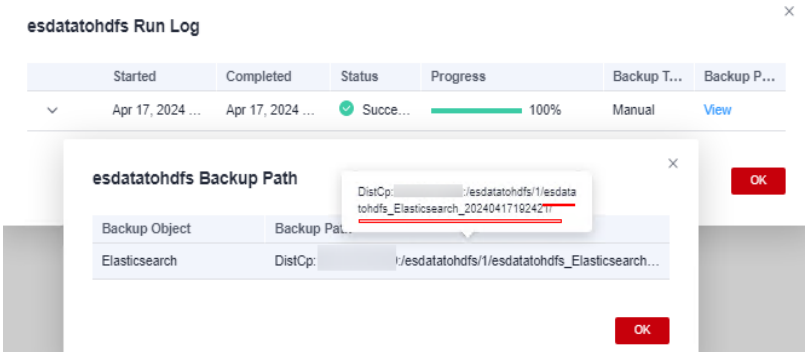
source bigdata_env

kinit *Service user* (Skip this step for normal clusters. Change the password upon the first authentication.)

Step 3 Create a snapshot repository.

NOTE

- **EsNode1_IP**: service IP address of any EsNode1 instance.
- **Snapshot name**: On the **Backup Management** page, click the created backup task and choose **More** > **View History**, and then click **View** in the **Backup Path** column to check the snapshot name.



- To obtain **core-site.xml** and **hdfs-site.xml** files, log in to FusionInsight Manager, choose **Cluster** > **Services** > **HDFS**. In the upper right corner, choose **More** > **Download Client**, set **Select Client Type** to **Configuration Files Only**, and click **OK** to download the configuration file package. Decompress the configuration file package and obtain the files from the **config** directory.
- The values of **conf.dfs.namenode.rpc-address.hacluster.19** and **conf.dfs.namenode.rpc-address.hacluster.20** parameters must be set based on the actual value of **dfs.ha.namenodes.hacluster**.
- If the HDFS cluster is in non-security mode, delete the **security.principal** and **hadoop.security.authentication** parameters.
- **--tlsv1.2** indicates that the TLSv1.2 protocol is used. This parameter can be omitted when you run the **curl** command. If the system displays a message indicating that TLSv1.2 is unavailable when this parameter is used, the curl client may not support TLSv1.2. You can run the **curl --help** command to query the TLS version. For example, if the TLSv1 protocol is used, set this parameter to **--tlsv1**.

```
curl -XPUT --tlsv1.2 --negotiate -k -u : "https://{EsNode1_IP}:24100/_snapshot/{Backup snapshot name}" -H 'Content-Type: application/json' -d'
{
"type": "hdfs",
"settings": {
"uri": "hdfs://hacluster/",        //core-site.xml fs.defaultFS
"path": "/test/backup",            //RemoteHDFS path to which data is backed up. On the Backup Management page, select the created backup task and choose More > View History to view the backup path.
"security.principal": "backup/manager@HADOOP.COM",    //Configure this parameter only when the HDFS component is in security mode. To obtain the domain name, log in to FusionInsight Manager, choose System > Permission > Domain and Mutual Trust, and view the value of Local Domain.
"hadoop.security.authentication":"kerberos",    //The authentication mode must be set to Kerberos. Set this parameter when the HDFS cluster is in security mode.
"load_defaults": true,
"concurrent_streams": 5,
"compress": "true",
"conf.hadoop.rpc.protection":"privacy",
"conf.dfs.nameservices":"hacluster",       //Set this parameter based on the value of dfs.nameservices in hdfs-site.xml.
"conf.dfs.ha.namenodes.hacluster": "19,20",//For details, see configurations of dfs.ha.namenodes.hacluster in hdfs-site.xml.
"conf.dfs.namenode.rpc-address.hacluster.19":"192.168.168.11:25000",//Set this parameter based on the dfs.namenode.rpc-address.hacluster.19 in hdfs-site.xml for configuration and replaces the host name with the corresponding service IP address.
"conf.dfs.namenode.rpc-address.hacluster.20":"192.168.168.22:25000",//Set this parameter based on the dfs.namenode.rpc-address.hacluster.20 in hdfs-site.xml for configuration and replaces the host name with the corresponding service IP address.
"conf.dfs.client.failover.proxy.provider.hacluster":"org.apache.hadoop.hdfs.server.namenode.ha.ConfiguredFail
```



```
overProxyProvider",
"max_snapshot_bytes_per_sec":"500mb",      //Specifies the snapshot rate limit of each node. The default value is 40 MB/s.
"max_restore_bytes_per_sec":"500mb"        //Specifies the throttle recovery rate of each node. The default value is 40 MB/s.
}
}'
```

The command output is as follows:

```
{"acknowledged":true}
```

Step 4 After the snapshot repository is created, you can run the following command to query the repository information:

curl -XGET --tlsv1.2 --negotiate -k -u : "https://{*EsNode1_IP*}*:24100*/**_snapshot/**{*Backup snapshot name*}**?pretty"**

----End

2.5.9 How Do I Configure the Environment When I Create a ClickHouse Backup Task on FusionInsight Manager and Set the Path Type to RemoteHDFS?

Question

How do I configure the environment when I create a ClickHouse backup task on FusionInsight Manager and set the path type to RemoteHDFS?

Answer

Step 1 Log in to FusionInsight Manager of the standby cluster.

Step 2 Choose **Cluster** > **Services** > **HDFS**. Click **More** and select **Download Client**. Set **Select Client Type** to **Configuration Files Only**, select **x86_64** for x86 or **aarch64** for ARM based on the type of the node where the client is to be installed, and click **OK**.

Step 3 After the client file package is generated, download the client to the local PC as prompted and decompress the package.

For example, if the client file package is **FusionInsight_Cluster_1_HDFS_Client.tar**, decompress it to obtain **FusionInsight_Cluster_1_HDFS_ClientConfig_ConfigFiles.tar**, and then decompress **FusionInsight_Cluster_1_HDFS_ClientConfig_ConfigFiles.tar** to the **D:\FusionInsight_Cluster_1_HDFS_ClientConfig_ConfigFiles** directory on the local PC. The directory name cannot contain spaces.

Step 4 Go to the **FusionInsight_Cluster_1_HDFS_ClientConfig_ConfigFiles** client directory and obtain the **hosts** file.

Step 5 Log in to FusionInsight Manager of the source cluster.

Step 6 Choose **Cluster** > **Services** > **ClickHouse**, click **Instance**, and view the instance IP address of **ClickHouseServer**.

Step 7 Log in to the host nodes of the ClickHouseServer instances as user **root** and check whether the **/etc/hosts** file contains the host information in **Step 4**. If not, add the host information in **Step 4** to the **/etc/hosts** file.

----End

2.5.10 Backing Up and Restoring ClickHouse Data Manually

2.5.10.1 Backing Up and Restoring Data Using a Data File

Scenarios

This section describes how to back up data by exporting ClickHouse data to a CSV file and restore data using the CSV file.

Prerequisites

- You have installed the ClickHouse client.
- You have created a user with related permissions on ClickHouse tables on Manager.
- You have prepared a server for backup.

Backing Up Data

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client installation directory.

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the current user. The current user must have the permission to create ClickHouse tables. If Kerberos authentication is disabled, skip this step.

kinit *Component service user*

Example: **kinit clickhouseuser**

Step 5 Run the ClickHouse client command to export the ClickHouse table data to be backed up to a specified directory.

clickhouse client --host *Host name or instance IP address* **--secure --port 21427 --query=**"*Table query statement*" > *Path of the exported CSV file*

The following shows an example of backing up data in the **test** table to the **default_test.csv** file on the ClickHouse instance **10.244.225.167**.

clickhouse client --host 10.244.225.167 --secure --port 21427 --query="select * from default.test FORMAT CSV" > /opt/clickhouse/default_test.csv

Step 6 Upload the exported CSV file to the backup server.

----End

Restoring Data

Step 1 Upload the backup data file on the backup server to the directory where the ClickHouse client is located.

For example, upload the **default_test.csv** backup file to the **/opt/clickhouse** directory.

Step 2 Log in to the node where the client is installed as the client installation user.

Step 3 Run the following command to go to the client installation directory.

cd /opt/hadoopclient

Step 4 Run the following command to configure environment variables:

source bigdata_env

Step 5 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the current user. The current user must have the permission to create ClickHouse tables. If Kerberos authentication is disabled, skip this step.

kinit *Component service user*

Example: **kinit clickhouseuser**

Step 6 Run the ClickHouse client command to log in to the ClickHouse cluster.

clickhouse client --host *Host name or instance IP address* **--secure --port 21427**

Step 7 Create a table with the format corresponding to the CSV file.

CREATE TABLE [IF NOT EXISTS] *[database_name.]table_name* **[ON CLUSTER** *Cluster name***]**

(

name1 **[***type1***] [DEFAULT|***materialized***|ALIAS** *expr1***],**

name2 **[***type2***] [DEFAULT|***materialized***|ALIAS** *expr2***],**

...

) *ENGINE = engine*

Step 8 Import the content in the backup file to the table created in **Step 7** to restore data.

clickhouse client --host *Host name or instance IP address* **--secure --port 21427 --query="insert into** *Table name* **FORMAT CSV" <** *CSV file path*

The following shows an example of restoring data from the **default_test.csv** backup file to the **test_cpy** table on the ClickHouse instance **10.244.225.167**.

clickhouse client --host 10.244.225.167 --secure --port 21427 --query="insert into default.test_cpy FORMAT CSV" < /opt/clickhouse/default_test.csv

----End

2.5.10.2 Backing Up and Restoring Data Using a CTAS Snapshot Table

Scenarios

This section describes how to back up table data by creating a CTAS snapshot table and importing data to the created table and how to restore data by inserting data in the CTAS snapshot table into the original table.

Prerequisites

- You have installed the ClickHouse client.
- You have created a user with related permissions on ClickHouse tables on Manager.

Backing Up Data

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client installation directory.

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the current user. The current user must have the permission to create ClickHouse tables. If Kerberos authentication is disabled, skip this step.

kinit *Component service user*

Example: **kinit clickhouseuser**

Step 5 Use the ClickHouse client for login.

clickhouse client --host *Host name or instance IP address* **--secure --port 21427**

Step 6 Create a CTAS backup table.

create table *CTAS table name* **as** *Name of the table to be backed up*;

For example, create a backup table **test_CTAS** based on the **test** table.

create table default.test_CTAS as default.test;

CAUTION

ReplicatedMergeTree tables do not support CTAS for backup.

Step 7 Import data in the original table to the CTAS backup table created in **Step 6**.

insert into table *CTAS table name* **select * from** *Name of the original table to be backed up*;

For example, back up data in the **test** table to the **test_CTAS** table.

insert into table test_CTAS select * from test;

----End

Restoring Data

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client installation directory.

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the current user. The current user must have the permission to create ClickHouse tables. If Kerberos authentication is disabled, skip this step.

kinit *Component service user*

Example: **kinit clickhouseuser**

Step 5 Use the ClickHouse client for login.

clickhouse client --host *Host name or instance IP address* **--secure --port 21427**

Step 6 Create a table to be restored.

create table *Name of the table to be restored* **as** *Name of the CTAS backup table*;

For example, create table **test** based on the **test_CTAS** backup table.

create table default.test as default.test_CTAS;

Step 7 Export data from the CTAS backup table to the table to be restored.

insert into table *Name of the table to be restored* **select * from** *Name of the CTAS backup table*;

For example, import table data backed up in the **test_CTAS** table to the **test** table.

insert into table test select * from test_CTAS ;

NOTE

You can perform remote backup and restoration using the remote query method.

Example: **insert into table default. test select * from remote (**'*remoteIp*'**,***Table name*,'*ClickHouse username*','*ClickHouse user password*'**);**

----End

2.5.10.3 Backing Up and Restoring Data Using FREEZE

Scenarios

This section describes how to back up table metadata and perform the **FREEZE** operation to generate a backup data file, and how to restore data by running the **ATTACH** statement to obtain the backup data file.

Prerequisites

- You have installed the ClickHouse client.
- You have created a user with related permissions on ClickHouse tables on Manager.

Backing Up Data

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to go to the client installation directory.

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the current user. The current user must have the permission to create ClickHouse tables. If Kerberos authentication is disabled, skip this step.

kinit *Component service user*

Example: **kinit clickhouseuser**

Step 5 Use the ClickHouse client for login.

clickhouse client --host *Host name or instance IP address* **--secure --port 21427**

Step 6 Perform merging on the table to be backed up.

OPTIMIZE TABLE *Table to be backed up* **FINAL;**

For example, perform merging on the **test** table.

OPTIMIZE TABLE default.test FINAL;

Step 7 Delete the shadow directory.

rm –rf /srv/BigData/clickHouse/data*N***/clickhouse/shadow**

Step 8 Run the **freeze** statement to generate a local backup file based on the table to be backed up.

alter table *Name of the original table to be backed up* **freeze;**

For example, run the **freeze** statement on the test table.

alter table default.test **freeze;**

Step 9 Query the data directory of the table to be backed up.

cd /srv/BigData/clickhouse/data*N*/**metadata/data/***Database name*

ls -ltr

For example, the following shows the metadata directory of the **test** table in the **default** database.

```
[root@kwephispra44948 default]# ls -ltr
total 12
lrwxrwxrwx 1 omm wheel 85 Jun 19 13:54 csv_tab001 -> /srv/BigData/clickhouse/data1/metadata/
store/b46/b46abb6b-6665-4cfd-80bb-0cecb87960df
lrwxrwxrwx 1 omm wheel 85 Jun 19 14:25 csv_tab0012 -> /srv/BigData/clickhouse/data1/metadata/
store/e70/e70845e8-33e2-428e-90af-e6c6e6942ab3
lrwxrwxrwx 1 omm wheel 85 Jun 21 14:45 test -> /srv/BigData/clickhouse/data1/metadata/store/c48/
c488eb86-ee85-4483-9d68-b708c8aa9af0
```

Step 10 Check whether the data file directory of the backup table is generated and obtain the table data file to be backed up. For example, **all_0_0_1** in the following figure.

cd /srv/BigData/clickhouse/data*N*/**clickhouse/shadow**/*x*/*Storage path of the table in* ***Step 9***

```
[root@kwephispra44948 c488eb86-ee85-4483-9d68-b708c8aa9af0]# cd /srv/BigData/clickhouse/data1/
clickhouse/shadow/1/store/c48/c488eb86-ee85-4483-9d68-b708c8aa9af0
[root@kwephispra44948 c488eb86-ee85-4483-9d68-b708c8aa9af0]# ls -ltr
total 4
drwxr-x--- 2 omm wheel 4096 Jun 21 14:27 all_0_0_1
```

Step 11 Obtain the metadata file of the table to be backed up from the ClickHouse metadata directory.

cd /srv/BigData/clickHouse/data1/metadata/metadata/*Database name*/

Run the **ll** command to query the file information. The following figure shows the table metadata in the **default** directory: *Backup table name.sql*

```
[omm@10-244-225-167 default]$ ll
total 20
-rw-r----- 1 omm wheel 144 Jun 17 12:55 test_cpy.sql
-rw-r----- 1 omm wheel 144 Jun 17 13:06 test_CTAS_Re.sql
-rw-r----- 1 omm wheel 144 Jun 17 13:02 test_CTAS.sql
-rw-r----- 1 omm wheel 146 Jun 17 12:43 test_dist.sql
-rw-r----- 1 omm wheel 144 Jun 17 13:08 test.sql
[omm@10-244-225-167 default]$ pwd
/srv/BigData/ClickHouse/metadata/metadata/default
```

Step 12 Archive the data file in **Step 10** and metadata file in **Step 11** to the backup server.

----End

Restoring Data

Step 1 You have obtained the data file and metadata file to be restored from the backup server.

Step 2 Obtain the metadata and table structure of the backup table from the metadata file.

Example:


```
[omm@10-244-225-167 default]$ cat test.sql
ATTACH TABLE _ UUID 'f43e486b-99ef-48ba-bbbc-68c07740cc21'
(
    `id` String
)
ENGINE = MergeTree
ORDER BY id
SETTINGS index_granularity = 8192
```

Step 3 Log in to the node where the client is installed as the client installation user.

Step 4 Run the following command to go to the client installation directory.

cd /opt/hadoopclient

Step 5 Run the following command to configure environment variables:

source bigdata_env

Step 6 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the current user. The current user must have the permission to create ClickHouse tables. If Kerberos authentication is disabled, skip this step.

kinit *Component service user*

Example: **kinit clickhouseuser**

Step 7 Create a table for restoration. Replace **ATTACH** in the metadata shown in **Step 2** with **CREATE** and replace the table name with the name of the table to be restored. Then, the statement can be used for table creation.

Example:

```
10-244-225-167 :) CREATE TABLE test
:-] (
:-]     `id` String
:-] )
:-] ENGINE = MergeTree
:-] ORDER BY id
:-] SETTINGS index_granularity = 8192

CREATE TABLE test
(
    `id` String
)
ENGINE = MergeTree
ORDER BY id
SETTINGS index_granularity = 8192

Query id: e9e712e6-32e6-4c51-8eb7-eee4693e3b27

Ok.

0 rows in set. Elapsed: 0.016 sec.
```

Step 8 Copy the backup data file to the **detached** directory in the table data directory and obtain the folder name. Run the following command to search for the data directory:

cd /srv/BigData/clickhouse/data*N*/**metadata/data/***Database name*

ls -ltr

The directory found by running this command is the mapped directory. Change **metadata** in the path to **clickhouse**, and then you will get the actual data directory.

For example, copy the backup table data file **all_1_1_0** to the **detached** directory and save the folder name.

```
[omm@10-244-225-167 detached]$ pwd
/srv/BigData/ClickHouse/data1/clickhouse/store/582/582f5ccc-23b8-488e-b499-67217a09dc97/detached
[omm@10-244-225-167 detached]$ ll
total 4
drwxr-x--- 2 omm wheel 4096 Jun 17 13:31 all_1_1_0
```

Step 9 Use the ClickHouse client for login.

clickhouse client --host *Host name or instance IP address* **--secure --port 21427**

Step 10 Restore data.

alter table *Name of the table to be restored* **attach part '***File name in **Step 8***';**

Example:

alter table default.test attach part 'all_1_1_0';

```
10-244-225-167 :) alter table default.test attach part 'all_1_1_0';

ALTER TABLE default.test
    ATTACH PART 'all_1_1_0'

Query id: 8d7ae9a0-3e06-4ab8-99bf-3dbbb439c6c0

Ok.

0 rows in set. Elapsed: 0.003 sec.

10-244-225-167 :) select * from default.test;

SELECT *
FROM default.test

Query id: ae53d2d6-c861-4012-87db-23153336ed88

┌─id─┐
│  1 │
│  2 │
│  3 │
└────┘

3 rows in set. Elapsed: 1.882 sec.
```

----End

3 Security Management

3.1 Security Overview

3.1.1 Security Overview

MRS uses LDAP as the account management system and uses Kerberos to perform security authentication on account information. The role definition information is stored in a relational database, and the mapping between roles and permissions is stored on each component. After a user is authenticated, MRS determines whether to authenticate the user based on the permission management configuration to ensure that the user has limited or all permissions on resources. If the user lack the permission for accessing cluster resources, the system administrator must grant the required permission to the user. Otherwise, the user fails to access the resources.

- MRS uses a unified user- and role-based authentication system as well as an account- and role-based access control (RBAC) model to centrally control user rights and batch manage user authorization.
- MRS supports the Hadoop Ranger authentication framework. Policy-based access control (PBAC) is used for permission management and implements fine-grained data access control for components such as HDFS, Hive, and HBase.

For details about Huawei Cloud Stack security policies, see "Maintenance Guide" > "O&M Guide" > "Security Management" in *Huawei Cloud Stack 8.3.1 Product Documentation* .

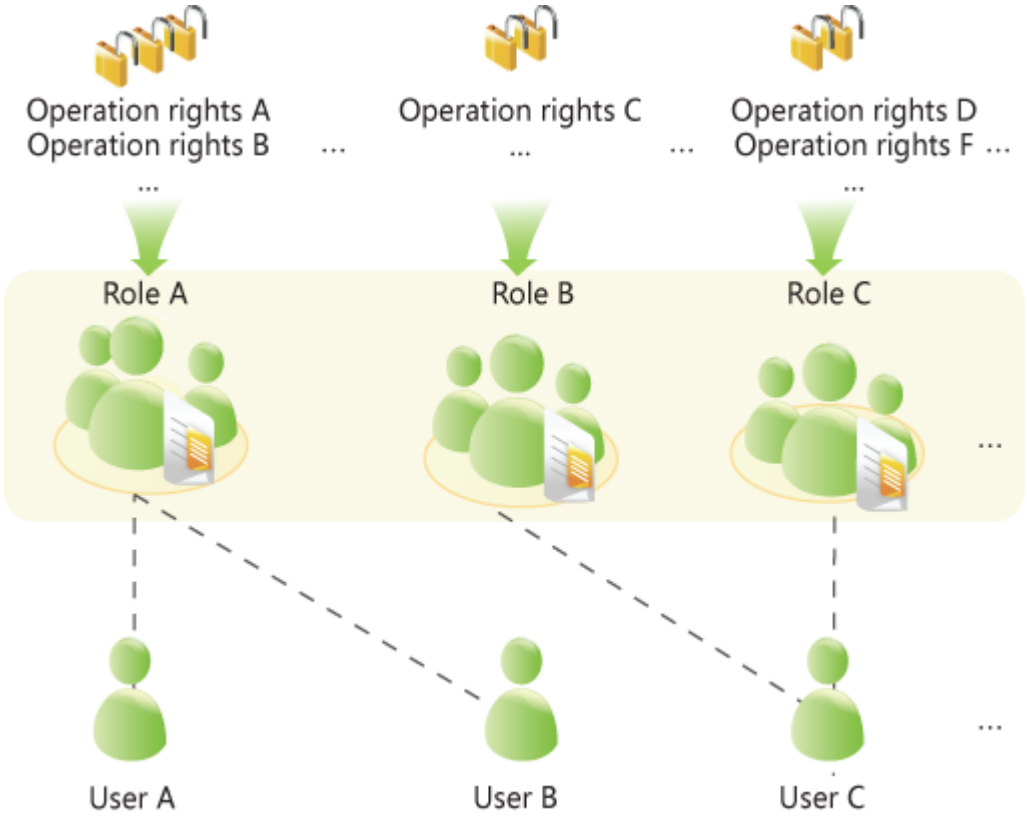
3.1.2 Right Model

Role-based Access Control

FusionInsight adopts the role-based access control (RBAC) mode to manage rights on the big data system. It integrates the right management functions of the components to centrally manage rights. Common users are shielded from internal right management details, and the right management operations are simplified for administrators, improving right management usability and user experience.

The right model of FusionInsight consists four parts, that is users, user groups, roles, and rights.

Figure 3-1 Right model



- **Right**

  Right, which is defined by components, allows users to access a certain resource of one component. Different components have different rights for their resources.

  For example:

  – HDFS provides read, write, and execute permissions on files.
  – HBase provides create, read, and write permissions on tables.

- **Role**

  Role is a collection of component rights. Each role can have multiple rights of multiple components. Different roles can have the rights of a resource of one component.

- **User group**

  User group is a collection of users. When a user group is bound to a role, users in this group obtain the rights defined by the role.

  Different user groups can be associated with the same role. A user group can also be associated with no role, and this user group does not have the rights of any component resources.

  NOTE

  In some components, the system grants related rights to specific user groups by default.

- **User**

  A user is a visitor to the system. Each user has the rights of the user group and role associated with the user. Users need to be added to the user group or associated with roles to obtain the corresponding rights.

Policy-based Access Control

The Ranger component uses policy-based access control (PBAC) to manage rights and implement fine-grained data access control on components such as HDFS, Hive, and HBase.

NOTE

The component supports only one right control mechanism. After the Ranger right control policy is enabled for the component, the right on the component in the role created on FusionInsight Manager becomes invalid (The ACL rules of HDFS and Yarn still take effect). You need to add a policy on the Ranger management page to grant rights on resources.

The Ranger right model consists of multiple right policies. A right policy consists of the following parts:

- Resource

  Resources are provided by components and can be accessed by users, such as HDFS files or folders, queues in Yarn, and databases, tables, and columns in Hive.

- User

  A User is a visitor to the system. The rights of each user are obtained based on the policy associated with the user. Information about users, user groups, and roles in the LDAP is periodically synchronized to the Ranger.

- Permission

  In a policy, you can configure various access conditions for resources, such as file read and write, permission conditions, rejection conditions, and exception conditions.

3.1.3 Right Mechanism

FusionInsight adopts the Lightweight Directory Access Protocol (LDAP) to store data of users and user groups. Information about role definitions is stored in the relational database and the mapping between roles and rights is saved in components.

FusionInsight uses Kerberos for unified authentication.

The verification process of user rights is as follows:

1. A client (a user terminal or FusionInsight component service) invokes the FusionInsight authentication interface.
2. FusionInsight uses the login username and password for Kerberos authentication.
3. If the authentication succeeds, the client sends a request for accessing the server (a FusionInsight component service).
4. The server finds the user group and role to which the login user belongs.

5. The server obtains all rights of the user group and the role.
6. The server checks whether the client has the right to access the resources it applies for.

Example (RBAC):

There are three files in HDFS, that is, fileA, fileB, and fileC.

- roleA has read and write right for fileA, and roleB has the read right for fileB.
- groupA is bound to roleA, and groupB is bound to roleB.
- userA belongs to groupA and roleB, and userB belongs to groupB.

When userA successfully logs in to the system and accesses the HDFS:

1. HDFS obtains the role (roleB) to which userA is bound.
2. HDFS also obtains the role (roleA) to which the user group of userA is bound.
3. In this case, userA has all the rights of roleA and roleB.
4. As a result, userA has read and write rights for fileA, has the read right on fileB, and has no right for fileC.

Similarly, when userB successfully logs in to the system and accesses the HDFS:

1. userB only has the rights of roleB.
2. As a result, userB has the read right on fileB, and has no rights for fileA and fileC.

3.1.4 Authentication Policies

The big data platform performs user identity authentication to prevent invalid users from accessing the cluster. The cluster provides authentication capabilities in both security mode and normal mode.

Security Mode

The clusters in security mode use the Kerberos authentication protocol for security authentication. The Kerberos protocol supports mutual authentication between clients and servers. This eliminates the risks incurred by sending user credentials over the network for simulated authentication. In clusters, KrbServer provides the Kerberos authentication support.

Kerberos user object

In the Kerberos protocol, each user object is a principal. A complete principal consists of username and domain name. In O&M or application development scenarios, the user identity must be verified before a client connects to a server. Users for O&M and service operations are classified into human-machine and machine-machine users. The password of human-machine users is manually configured, while the password of machine-machine users is generated by the system randomly.

Kerberos authentication

Kerberos supports password and keytab authentication. The validity period of authentication is 24 hours by default.

- Password authentication: User identity is verified by entering the correct password. This mode mainly used in O&M scenarios where human-machine users are used. The configuration command is **kinit** *Username*.
- Keytab authentication: Keytab files contain users' principal and encrypted credential information. When keytab files are used for authentication, the system automatically uses encrypted credential information to perform authentication and the user password does not need to be entered. This mode is mainly used in component application development scenarios where machine-machine users are used. Keytab authentication can also be configured using the **kinit** command.

Normal Mode

Different components in a normal cluster use the native open-source authentication mode and do not support the **kinit** authentication command. FusionInsight Manager (including DBService, KrbServer, and LdapServer) uses the username and password for authentication. **Table 3-1** lists the authentication modes used by components.

Table 3-1 Component authentication modes

Service	Authentication Mode
IoTDB	Simple authentication
CDL	No authentication
ClickHouse	Simple authentication
Elasticsearch	Client: simple authentication
Flume	No authentication
FTP-Server	Username and password authentication
HBase	• Web UI: no authentication • Client: simple authentication
HDFS	• Web UI: no authentication • Client: simple authentication
HetuEngine	• Web UI: no authentication • Client: no authentication
Hive	Simple authentication
Hue	Username and password authentication
Kafka	No authentication
Loader	• Web UI: username and password authentication • Client: no authentication
MapReduce	• Web UI: no authentication • Client: no authentication

Service	Authentication Mode
Oozie	• Web UI: username and password authentication • Client: simple authentication
Redis	No authentication
Solr	No authentication
Spark	• Web UI: no authentication • Client: simple authentication
Yarn	• Web UI: no authentication • Client: simple authentication
ZooKeeper	Simple authentication
MOTService	Username and password authentication
Containers	No authentication
RTDService	Username and password authentication
Guardian	No authentication
MemArtsCC	No authentication

The authentication modes are as follows:

- Simple authentication: When the client connects to the server, the client automatically authenticates the user (for example, the OS user **root** or **omm**) by default. The authentication is imperceptible to the administrator or service user, which does not require **kinit**.
- Username and password authentication: Use the username and password of human-machine users in the cluster for authentication.
- No authentication: Any user can access the server by default.

3.1.5 Permission Verification Policies

Security Mode

After a user is authenticated by the big data platform, the system determines whether to verify the user's permission based on the actual permission management configuration to ensure that the user has limited or all permission on resources. If the user does not have the permission for accessing cluster resources, the system administrator must grant the required permission to the user. Otherwise, the user fails to access the resources. The cluster provides permission verification capabilities in both security mode and normal mode. The specific permission items of the components are the same in the two modes.

By default, the Ranger service is installed and Ranger authentication is enabled for a newly installed cluster in security mode. You can set fine-grained security access policies for accessing component resources through the permission plug-in of the

component. If Ranger authentication is not required, administrators can manually disable it on the service page. After Ranger authentication is disabled, the system continues to perform permission control based on the role model of FusionInsight Manager when accessing component resources.

In a cluster in security mode, the following components support Ranger authentication: HDFS, Yarn, Kafka, Hive, HBase, Elasticsearch, HetuEngine, CDL, and Spark.

For a cluster upgraded from an earlier version, Ranger authentication is not used by default when users access component resources. The administrator can manually enable Ranger authentication after installing Ranger.

By default, all components in the cluster of the security edition authenticate access. The authentication function cannot be disabled.

Normal Mode

Different components in a normal cluster use their own native open-source authentication behavior. **Table 3-2** lists detailed permission verification modes.

In a normal cluster, Ranger authentication can be used to perform permission control by OS user on components such as HBase, HDFS, Hive, Spark, and Yarn.

Table 3-2 Component permission verification modes in normal clusters

Service	Permission Verification	Permission Verification Enabling and Disabling
IoTDB	Required	Not supported
ClickHouse	Required	Not supported
Flume	Not required	Not supported
FTP-Server	Depends on HDFS permission verification	Not supported
HBase	Not required	Supported
HDFS	Required	Supported
HetuEngine	Not required	Not supported
Hive	Not required	Not supported
Hue	Not required	Not supported
Kafka	Not required	Not supported
Loader	Not required	Not supported
MapReduce	Not required	Not supported
Oozie	Required	Not supported
Redis	Not required	Not supported
Solr	Not required	Not supported

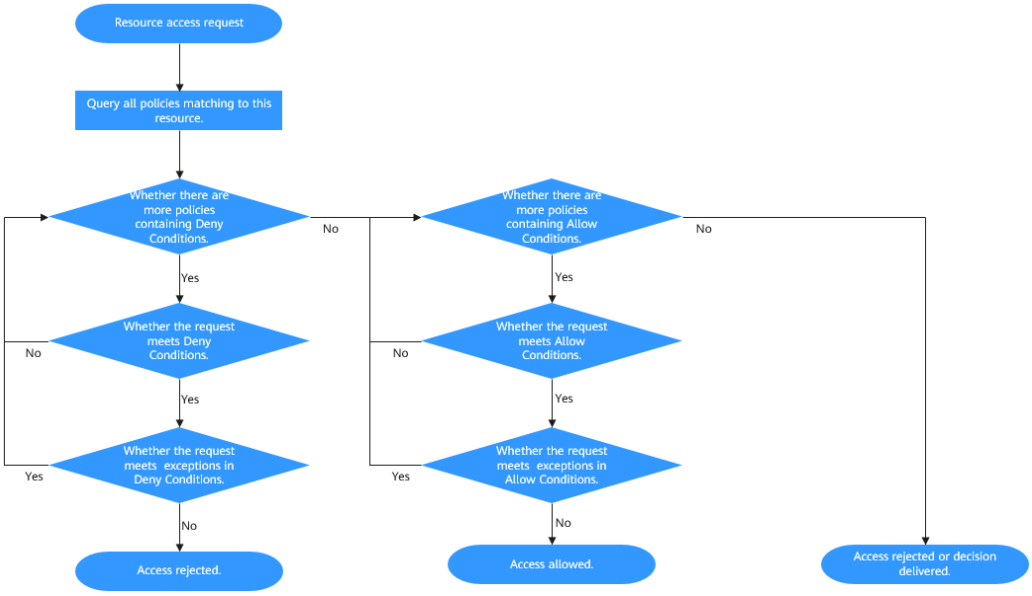
Service	Permission Verification	Permission Verification Enabling and Disabling
Spark	Not required	Not supported
Yarn	Not required	Supported
ZooKeeper	Required	Supported
CDL	Not required	Not supported
Elasticsearch	Not required	Not supported
Containers	Not required	Not supported
RTDService	Required	Not supported
MOTService	Required	Supported
Doris	Required	Not supported
Guardian	Not required	Not supported
MemArtsCC	Not required	Not supported

Condition Priorities of the Ranger Permission Policy

When configuring a permission policy for a resource, you can configure Allow Conditions, Exclude from Allow Conditions, Deny Conditions, and Exclude from Deny Conditions for the resource, to meet unexpected requirements in different scenarios.

The priorities of different conditions are listed in descending order: Exclude from Deny Conditions > Deny Conditions > Exclude from Allow Conditions > Allow Conditions

The following figure shows the process of determining condition priorities. If the component resource request does not match the permission policy in Ranger, the system rejects the access by default. However, for HDFS and Yarn, the system delivers the decision to the access control layer of the component for determination.



For example, if you want to grant the read and write permissions of the **FileA** folder to the **groupA** user group, but the user in the group is not **UserA**, you can add an allowed condition and an exception condition.

3.1.6 Weak Password Dictionary

Scenario

FusionInsight Manager provides the unified weak password verification function and weak password dictionary maintenance function. When a user sets a password, the system verifies the password and forbids the use of character strings in the weak password dictionary. The specific operations are as follows:

- Create a user.
- Forcibly change the password upon the first login.
- Initialize the passwords of other users as an administrator.
- Change the password as a logged-in user.
- Change the OMM database password.
- Change the password of an LDAP administrator account.
- Choose **System** > **OMS** and change the GaussDB password.
- Run the **kpasswd** command to change the password of the Kerberos administrator or change the password of the OMS Kerberos administrator.
- Run the **kpasswd** command to change the password of a system user on a client.

Weak Password Verification

When a user sets a password, the system checks whether the password is a weak password. Character strings in the weak password dictionary cannot be used. When a weak password is used, the system asks the user to change the password. The system displays a message indicating that the new password is a weak or common password and has security risks.

Weak Password Dictionary Update

FusionInsight Manager updates the weak password dictionary during an upgrade or patch installation.

3.1.7 Default Permission Information

Role

Default Role	Description
Manager_administrator	Manager administrator who has all permissions for Manager. Manager administrators can create first-level tenants, create and modify user groups, and specify user permissions.
Manager_operator	Manager operator who has all the permissions on the **Homepage**, **Cluster**, **Hosts**, and **O&M** tab pages.
Manager_auditor	Manager auditor who has all permissions on the **Audit** tab page. Manager auditors can view and manage Manager system audit logs.
Manager_viewer	Manager viewer who has the permission to view information about **Homepage**, **Cluster**, **Hosts**, **Alarm**, **Events**, and **System > Permission**, and download clients.
Manager_tenant	Manager tenant administrator. This role can create and manage sub-tenants for the non-leaf tenants to which the current user belongs. It has the permission to view alarms and events on **O&M > Alarm**.
System_administrator	System administrator, this role has Manager system administrator rights and all services administrator rights.
default	This role is the default role created for the **default** tenant. It has the management permissions on the Yarn component and the default queue. The default role of the default tenant that is not the first cluster to be installed is **c***<cluster ID>*_**default**.
Manager_administrator_180	FusionInsight Manager system administrator group. Internal system user group, which is used only between components.
Manager_auditor_181	FusionInsight Manager system auditor group. Internal system user group, which is used only between components.

Default Role	Description
Manager_operator_182	FusionInsight Manager system operator group. Internal system user group, which is used only between components.
Manager_viewer_183	FusionInsight Manager system viewer group. Internal system user group, which is used only between components.
System_administrator_186	System administrator group. Internal system user group, which is used only between components.
Manager_tenant_187	Tenant system user group. Internal system user group, which is used only between components.
default_1000	This group is created for tenant. Internal system user group, which is used only between components.

User group

Type	Default User Group	Description
Default cluster user groups	cdl	Common user group of CDL. Users in this group can create and query CDL jobs.
	cdladmin	CDL administrator group. Only users in this group can access CDL APIs.
	Elasticsearch	Users added to this user group can use Elasticsearch.
	graphbaseadmin	GraphBase administrator group. Users added to this user group will have the administrator rights of GraphBase and GraphServer.
	graphbasedeveloper	GraphBase developer group. Users added to this user group will have the developer rights of GraphBase and GraphServer.
	graphbaseoperator	GraphBase operator group. Users in this group have the permission to query data on the GraphServer web UI.
	hadoop	Users added to this group are granted the permission to submit all Yarn queue tasks.
	hadoopmanager	Users added to this user group can have the O&M manager rights of HDFS and Yarn. The O&M manager of HDFS can access the NameNode WebUI and perform active to standby switchover manually. The O&M manager of Yarn can access the ResourceManager WebUI, operate NodeManager nodes, refresh queues, and set node labels, but cannot submit tasks.

Type	Default User Group	Description
	hetuadmin	HetuEngine administrator group. Users in this group have the permission to perform operations on HSConsole.
	hetuuser	User group to which the users need to be added to obtain the SQL execution permission
	hive/ hive1/ hive2/ hive3/ hive4	Common user group. Hive/Hive1/Hive2/Hive3/Hive4 users must be in this user group.
	iotdbgroup	Users added to this user group have the administrator rights of the IoTDB component.
	kafka	Kafka common user group. A user in this group can access a topic only when a user in the kafkaadmin group grants the read and write permission of the topic to the user.
	kafkaadmin	Kafka administrator group. Users in this group have the rights to create, delete, authorize, read, and write all topics.
	kafkasuperuser	Topic read/write user group of Kafka. Users added to this group have the read and write permissions on all topics.
	kafkaui	Kafka UI user group. Users in this group have the permission to view Kafka UI.
	kmsadmin	After a user is added to the user group, the read permission on all keys in the KMS can be obtained.
	msadmin	Users added to this user group have the administrator rights of Metastore.
	rkmsadmin	User group for RangerKMS permission management. If the key management permission is required, add the user to this group.
	solr	Users added to this user group can use Solr.
	supergroup	Users added to this user group can have the administrator rights of HBase, HDFS, Solr, Redis, and Yarn and can use Hive.
	yarnviewgroup	Indicates the read-only user group of the Yarn task. Users in this user group can have the view permission on Yarn and MapReduce tasks.
	check_sec_ldap	Perform internal test on the active LDAP to see whether it works properly. This user group is generated randomly in a test and automatically deleted after the test is complete. Internal system user group, which is used only between components.

Type	Default User Group	Description
	compcommon	System internal group for accessing cluster system resources. All system users and system running users are added to this user group by default.
Default cluster user groups	lakesearch group	Users added to this user group have the administrator rights of the LakeSearch component.
OS user groups	wheel	Primary group of the FusionInsight internal running user omm.
	ficommon	System common group that corresponds to **compcommon** for accessing cluster common resource files stored in the OS.

NOTE

If the current cluster is not the cluster that is installed for the first time in FusionInsight Manager, the default user group name of all components except Manager in the cluster is **c***<cluster ID>_ default user group name*, for example, **c2_hadoop**.

User

For details, see ***Huawei Cloud Stack 8.3.1 Account List*** .

Service-related User Security Parameters

- **FTP-Server**
  - The **ftp-group** parameter specifies the user group to which common users who are allowed to connect to the FTP server belong. If the users are not added to the corresponding user group, they cannot connect to the FTP server. The default value is **hadoop**.
  - The **ftp-admin-group** parameter specifies the user group to which the administrator of the FTP server belongs. If the administrator is not added to the corresponding user group, the administrator cannot operate directories and files of other users. The default value is **supergroup**.
- **HDFS**

  The **dfs.permissions.superusergroup** parameter specifies the administrator group with the highest permission on the HDFS. The default value is **supergroup**.
- **Spark and Corresponding Multi-Instances**

  The **spark.admin.acls** parameter indicates the Spark administrator list. Members in the list have the permission to manage all Spark tasks. If a user is

not added to the list, the user cannot manage all Spark tasks. The default value is **admin**.

3.2 Account Management

3.2.1 Account List

For details about accounts on the MRS console and accounts of nodes and components in the MRS cluster, see ***Huawei Cloud Stack 8.3.1 Account List*** .

3.2.2 Account Security Settings

3.2.2.1 Unlocking LDAP Users and Management Accounts

Scenario

If the LDAP user **cn=pg_search_dn,ou=Users,dc=hadoop,dc=com** and LDAP management accounts **cn=krbkdc,ou=Users,dc=hadoop,dc=com** and **cn=krbadmin,ou=Users,dc=hadoop,dc=com** are locked, the administrator must unlock these accounts.

NOTE

If you input an incorrect password for the LDAP user or management account for five consecutive times, the LDAP user or management account is locked. The account is automatically unlocked after 5 minutes.

Procedure

Step 1 Log in to the active management node as user **omm**.

Step 2 Run the following command to go to the related directory:

cd ${BIGDATA_HOME}/om-server/om/ldapserver/ldapserver/local/script

Step 3 Run the following command to unlock the LDAP user or management account:

./ldapserver_unlockUsers.sh *USER_NAME*

In the command, *USER_NAME* indicates the name of the user to be unlocked.

For example, to unlock the LDAP management **account cn=krbkdc,ou=Users,dc=hadoop,dc=com**, run the following command:

./ldapserver_unlockUsers.sh krbkdc

After the script is executed, enter the password of user **krbkdc** next to **ROOT_DN_PASSWORD**. For details about the default password, see ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator. If the following command output is displayed, the account is successfully unlocked:

```
Unlock user krbkdc successfully.
```

----End

3.2.2.2 Internal an Internal System User

Scenario

If the service is abnormal, the internal user of the system may be locked. Unlock the user promptly, or the cluster cannot run properly. For the list of system internal users, see ***Huawei Cloud Stack 8.3.1 Account List*** . System internal users cannot be unlocked using FusionInsight Manager.

Prerequisites

You have obtained the default password of the LDAP administrator **cn=root,dc=hadoop,dc=com** by referring to ***Huawei Cloud Stack 8.3.1 Account List*** .

Procedure

Step 1 Use the following method to confirm whether the internal system username is locked:

1. OLdap port number obtaining method:
   a. Log in to FusionInsight Manager, choose **System** > **OMS** > **oldap** > **Modify Configuration**.
   b. The **LDAP Listening Port** parameter value is **oldap port**.
2. Domain name obtaining method:
   a. Log in to FusionInsight Manager, choose **System** > **Permission** > **Domain and Mutual Trust**.
   b. The **Local Domain** parameter value is the domain name.

      For example, the domain name of the current system is **9427068F-6EFA-4833-B43E-60CB641E5B6C.COM**.
3. Run the following command on each node in the cluster as user **omm** to query the number of password authentication failures:

   ldapsearch -H ldaps://*OMS Floating IP Address***:***OLdap port* **-LLL -x -D cn=root,dc=hadoop,dc=com -b krbPrincipalName=***Internal system username***@***Domain name***,cn=***Domain name***,cn=krbcontainer,dc=hadoop,dc=com -w** *Password of LDAP administrator* **-e ppolicy | grep krbLoginFailedCount**

   For example, run the following command to check the number of password authentication failures for user **oms/manager**:

   ldapsearch -H ldaps://10.5.146.118:21750 -LLL -x -D cn=root,dc=hadoop,dc=com -b krbPrincipalName=oms/manager@9427068F-6EFA-4833-B43E-60CB641E5B6C.COM,cn=9427068F-6EFA-4833-B43E-60CB641E5B6C.COM,cn=krbcontainer,dc=hadoop,dc=com -w *Password of user cn=root,dc=hadoop,dc=com* **-e ppolicy | grep krbLoginFailedCount**

   ```
   krbLoginFailedCount: 5
   ```
4. Log in to FusionInsight Manager, choose **System** > **Permission** > **Security Policy** > **Password Policy**.

5. Check the value of the **Password Retries** parameter. If the value is less than or equal to the value of **krbLoginFailedCount**, the user is locked.

   NOTE

   You can also check whether internal users are locked by viewing operations logs.

Step 2 Log in to the active management node as user **omm** and run the following command to unlock the user:

sh ${BIGDATA_HOME}/om-server/om/share/om/acs/config/unlockuser.sh --userName *Internal system username*

Example: **sh ${BIGDATA_HOME}/om-server/om/share/om/acs/config/unlockuser.sh --userName oms/manager**

----End

3.2.2.3 Enabling and Disabling Permission Verification on Cluster Components

Scenario

HDFS and ZooKeeper verify the permission of users who attempt to access the services in both security and normal clusters by default. Users without related permission cannot access resources in HDFS and ZooKeeper. When the cluster is deployed in normal mode, HBase and Yarn do not verify the permission of users who attempt to access the services by default. All users can access resources in HBase and Yarn.

Based on actual service requirements, administrators can enable permission verification on HBase and Yarn or disable permission verification on HDFS and ZooKeeper in normal clusters.

Impact on the System

After the enabling and disabling operations, the service configuration will expire. You need to restart the corresponding service for the configuration to take effect.

Enabling Permission Verification on HBase

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > **Services** > **HBase** and click **Configurations**.

Step 3 Click **All Configurations**.

Step 4 Search for parameters **hbase.coprocessor.region.classes**, **hbase.coprocessor.master.classes**, and **hbase.coprocessor.regionserver.classes**.

Add the coprocessor parameter **org.apache.hadoop.hbase.security.access.AccessController** to the end of the values of the preceding parameters, and use a comma (,) to separate the values from those of the original coprocessors.

Step 5 Click **Save**, click **OK**, and wait for message "Operation successful" to display.

----End

Disabling Permission Verification on HBase

NOTE

After HBase permission verification is disabled, the existing permission data will be retained. If you want to delete permission information, disable permission verification, enter the HBase shell, and delete table **hbase:acl**.

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > **Services** > **HBase** and click **Configurations**.

Step 3 Click **All Configurations**.

Step 4 Search for parameters **hbase.coprocessor.region.classes**, **hbase.coprocessor.master.classes**, and **hbase.coprocessor.regionserver.classes**.

Delete the coprocessor parameter **org.apache.hadoop.hbase.security.access.AccessController**.

Step 5 Click **Save**, click **OK**, and wait for message "Operation successful" to display.

----End

Disabling Permission Verification on HDFS

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > **Services** > **HDFS** and click **Configurations**.

Step 3 Click **All Configurations**.

Step 4 Search for parameters **dfs.namenode.acls.enabled** and **dfs.permissions.enabled**.

- **dfs.namenode.acls.enabled** indicates whether to enable HDFS ACL. The default value is **true**, indicating that the ACL is enabled. Change the value to **false**.
- **dfs.permissions.enabled** indicates whether to enable permission check for HDFS. The default value is **true**, indicating that permission check is enabled. Change the value to **false**. After the modification, the owner, owner group, and permission of the directories and files in HDFS remain unchanged.

Step 5 Click **Save**, click **OK**, and wait for message "Operation successful" to display.

----End

Enabling Permission Verification on Yarn

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > **Services** > **Yarn** and click **Configurations**.

Step 3 Click **All Configurations**.

Step 4 Search for parameter **yarn.acl.enable**.

yarn.acl.enable indicates whether to enable the permission check for Yarn.

- In normal clusters, the value is set to **false** by default to disable permission check. To enable permission check, change the value to **true**.

- In security clusters, the value is set to **true** by default to enable authentication.

Step 5 Click **Save**, click **OK**, and wait for message "Operation successful" to display.

----End

Disabling Permission Verification on ZooKeeper

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > **Services** > **ZooKeeper** and click **Configurations**.

Step 3 Click **All Configurations**.

Step 4 Search for parameter **skipACL**.

skipACL indicates whether to skip the ZooKeeper permission check. The default value is **no**, indicating that permission check is enabled. Change the value to **yes**.

Step 5 Click **Save**, click **OK**, and wait for message "Operation successful" to display.

----End

3.2.2.4 Logging In to a Non-Cluster Node Using a Cluster User in Normal Mode

Scenario

When the cluster is installed in normal mode, the component clients do not support security authentication and cannot use the **kinit** command. Therefore, nodes outside the cluster cannot use users in the cluster by default. This may result in a user authentication failure when one of these nodes access a component server.

The node administrator can configure a user who has the same name as that of a user for a node outside the cluster, allow the user to log in to the node using the SSH protocol, and connect to the servers of components in the cluster by using the user who logs in to the OS.

Prerequisites

- Nodes outside the cluster can connect to the service plane of the cluster.
- The KrbServer service of the cluster is running properly.
- You have obtained the password of user **root** of the node outside the cluster.
- A human-machine user has been planned and added to the cluster, and you have obtained the authentication credential file. For details, see **Creating a User** and **Exporting an Authentication Credential File**.

Procedure

Step 1 Log in to the node where a user is to be added as user **root**.

Step 2 Run the following command:

rpm -qa | grep pam and **rpm -qa| grep krb5-client**

The following RPM packages are displayed:

```
pam_krb5-32bit-2.3.1-47.12.1
pam-modules-32bit-11-1.22.1
yast2-pam-2.17.3-0.5.211
pam-32bit-1.1.5-0.10.17
pam_mount-32bit-0.47-13.16.1
pam-config-0.79-2.5.58
pam_krb5-2.3.1-47.12.1
pam-doc-1.1.5-0.10.17
pam-modules-11-1.22.1
pam_mount-0.47-13.16.1
pam_ldap-184-147.20
pam-1.1.5-0.10.17
krb5-client-1.6.3
```

Step 3 Check whether the RPM packages in the list are installed in the OS.

- If yes, go to **Step 5**.
- If no, go to **Step 4**.

Step 4 Obtain the lacked RPM packages from the OS image, upload the files to the current directory, and run the following command to install the RPM package:

rpm -ivh *.rpm

NOTE

The RPM packages to be installed may bring security risks. The risks that may be brought by the installation of these RPM packages must be taken into consideration during OS hardening.

After the RPM packages are installed, go to **Step 5**.

Step 5 Run the following command to configure Kerberos authentication on PAM:

pam-config --add --krb5

NOTE

If you need to cancel Kerberos authentication and system user login on a non-cluster node, run the **pam-config --delete --krb5** command as user **root**.

Step 6 Decompress the authentication credential file to obtain **krb5.conf**, use WinSCP to upload this configuration file to the **/etc** directory on the node outside the cluster, and run the following command to configure related permission to enable other users to access the file, such as permission **604**:

chmod 604 /etc/krb5.conf

Step 7 Run the following command in the connection session as user **root** to add the corresponding OS user to the human-machine user, and specify **root** as the primary group.

The OS user password is the same as the initial password when the human-machine user is created on Manager.

useradd *User name* **-m -d /home/admin_test -g root -s /bin/bash**

For example, if the name of the human-machine user is **admin_test**, run the following command:

useradd admin_test -m -d /home/admin_test -g root -s /bin/bash

NOTE

When you use the newly added OS user to log in to the node by using the SSH protocol for the first time, the system prompts that the password has expired after you enter the user password, and the system prompts that the password needs to be changed after you enter the user password again. You need to enter a new password that meets the password complexity requirements of both the node OS and the cluster.

----End

3.2.3 Password Changing

3.2.3.1 Changing the Password for a System User

3.2.3.1.1 Changing the Password for User admin

Scenario

User **admin** is the system administrator account of FusionInsight Manager. You are advised to periodically change the password on FusionInsight Manager to improve system security.

Procedure

Step 1 Log in to FusionInsight Manager.

User **admin** is required for login.

Step 2 Move the cursor to **Hello, admin** in the upper right corner of the page.

In the displayed menu, click **Change Password**.

Step 3 Set **Old Password**, **New Password**, and **Confirm Password**, and click **OK**.

The password must meet the following complexity requirements by default:

- The password contains 8 to 64 characters.
- The password contains at least four types of the following characters: Uppercase letters, lowercase letters, digits, spaces, and special characters which can only be ~`!?,.;-_'(){}[]/<>@#$%^&*+|\=.
- The password cannot be the same as the username or the username spelled backwards.
- The password cannot be a common easily-cracked passwords.
- The password cannot be the same as the password used in the last *N* times. *N* indicates the value of **Repetition Rule** in **Configuring Password Policies**.

----End

3.2.3.1.2 Changing the Password for an OS User

Scenario

During FusionInsight Manager installation, the system automatically creates user **omm** and **ommdba** on each node in the cluster. Periodically change the login

passwords of the OS users **omm** and **ommdba** of the cluster node to improve the system O&M security.

The passwords of users **omm** and **ommdba** of the nodes can be different.

Prerequisites

- You have obtained the IP address of the node where the passwords of users **omm** and **ommdba** are to be changed.
- You have obtained the password of user **root** before changing the passwords of users **omm** and **ommdba**.

Changing the Password of an OS User

Step 1 Log in to the node where the password is to be changed as user **root**.

Step 2 Run the following command to change the user password:

passwd *ommdba*

Red Hat system displays the following information:

```
Changing password for user ommdba.
New password:
```

Step 3 Enter a new password. The policy for changing the password of an OS user varies according to the OS that is actually used.

```
Retype New Password:
Password changed.
```

----End

3.2.3.2 Changing the Password for a System Internal User

3.2.3.2.1 Changing the Password for the Kerberos Administrator

Scenario

It is recommended that the administrator periodically change the password of Kerberos administrator **kadmin** to improve the system O&M security.

If the user password is changed, the OMS Kerberos administrator password is changed as well.

Prerequisites

You have installed the client on any node in the cluster and obtained the IP address of the node.

Procedure

Step 1 Log in to the node where the client is installed as user **root**.

Step 2 Run the following command to go to the client directory, for example, **/opt/hadoopclient**:

cd /opt/hadoopclient

Step 3 Run the following command to set environment variables:

source bigdata_env

Step 4 Run the following command to change the password for **kadmin/admin**. The password changing takes effect on all servers. Keep the password secure because it cannot be retrieved once lost.

kpasswd kadmin/admin

The password must meet the following complexity requirements by default:

- The password contains at least 8 characters.
- The password contains at least four types of the following characters: Uppercase letters, lowercase letters, digits, spaces, and special characters which can only be ~`!?,.;-_'(){}[]/<>@#$%^&*+|\=.
- The password cannot be the same as the username or the username spelled backwards.
- The password cannot be a common easily-cracked passwords.
- The password cannot be the same as the password used in the last *N* times. *N* indicates the value of **Repetition Rule** in **Configuring Password Policies**.

NOTE

If you fail to change the password, see **How Do I Solve the Error That Occurs During the kpasswd Command Execution?**.

----End

3.2.3.2.2 Changing the Password for the OMS Kerberos Administrator

Scenario

It is recommended that the administrator periodically change the password of OMS Kerberos administrator **kadmin** to improve the system O&M security.

If the user password is changed, the Kerberos administrator password is changed as well.

Procedure

Step 1 Log in to any management node in the cluster as user **omm**.

Step 2 Run the following command to go to the related directory:

cd ${BIGDATA_HOME}/om-server/om/meta-0.0.1-SNAPSHOT/kerberos/scripts

Step 3 Run the following command to set environment variables:

source component_env

Step 4 Run the following command to change the password for **kadmin/admin**. The password changing takes effect on all servers. Keep the password secure because it cannot be retrieved once lost.

kpasswd kadmin/admin

The password must meet the following complexity requirements by default:

- The password contains at least 8 characters.
- The password contains at least four types of the following characters: uppercase letters, lowercase letters, digits, and special characters can only be ~`!?,.:;-_'(){}[]/<>@#$%^&*+|\=.
- The password cannot be the same as the username or the username spelled backwards.
- The password cannot be a common easily-cracked passwords.
- The password cannot be the same as the password used in the last *N* times. *N* indicates the value of **Repetition Rule** in **Configuring Password Policies**.

NOTE

If you fail to change the password, see **How Do I Solve the Error That Occurs During the kpasswd Command Execution?**.

----End

3.2.3.2.3 Changing the Password for a Component Running User

Scenario

It is recommended that the administrator periodically change the password for each component running user to improve the system O&M security.

Component running users can be classified into the following two types depending on whether their initial passwords are randomly generated by the system:

- If the initial password of a component running user is randomly generated by the system, the user is of the machine-machine type.
- If the initial password of a component running user is not randomly generated by the system, the user is of the human-machine type.

Impact on the System

If the initial password is randomly generated by the system, the cluster needs to be restarted for the password changing to take effect. Services are unavailable during the restart.

Prerequisites

You have installed the client on any node in the cluster and obtained the IP address of the node.

Procedure

Step 1 Log in to the node where the client is installed as the client installation user

Step 2 Run the following command to switch to the client directory, for example, **/opt/hadoopclient**:

cd /opt/hadoopclient

Step 3 Run the following command to set environment variables:

source bigdata_env

Step 4 Run the following command and enter the password of user **kadmin/admin** to log in to the **kadmin** console:

kadmin -p kadmin/admin

NOTE

The default password of user **kadmin/admin** can be obtained by referring to ***Huawei Cloud Stack 8.3.1 Account List*** or contacting the system administrator. The password will expire upon your first login. Change the password as prompted. Keep the password secure because it cannot be retrieved once lost.

Step 5 Run the following command to change the password of an internal component running user.

cpw *Internal system username*

Example: **cpw hdfs**

User **hdfs** is an example. Replace it with the actual username.

The password must meet the following complexity requirements by default:

- Contains at least 8 characters.
- Contains at least four types of the following: uppercase letters, lowercase letters, numbers, spaces, and special characters (~`!?,.;-_'(){}[]/<>@#$%^&*+|\=).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be a common easily-cracked password.
- Cannot be the same as the password used in the latest *N* times. *N* indicates the value of **Repetition Rule** configured in **Configuring Password Policies**. This policy applies only to human-machine accounts.

NOTE

Run the following command to check user information:

getprinc *Internal system username*

Example: **getprinc hdfs**

Step 6 Determine the type of the user whose password needs to be changed.

- If the user is a machine-machine user, go to **Step 7**.
- If the user is a human-machine user, the password is changed successfully and no further action is required.

Step 7 Log in to FusionInsight Manager.

Step 8 In the upper right corner of **Homepage**, click **More** and select **Restart**.

Step 9 In the displayed dialog box, enter the password of the current login user and click **OK**.

Step 10 In the displayed restart confirmation dialog box, click **OK**.

Step 11 Wait for message "Operation successful" to display.

----End

3.2.3.3 Changing the Password for a Database User

3.2.3.3.1 Changing the Password of the OMS Database Administrator

Scenario

It is recommended that the administrator periodically change the password of the OMS database administrator to improve the system O&M security.

Procedure

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

NOTE

The password of user **ommdba** cannot be changed on the standby management node. Otherwise, the cluster may not work properly. Change the password on the active management node only.

Step 2 Run the following command to switch to another user:

su - omm

Step 3 Run the following command to switch the directory:

cd $OMS_RUN_PATH/tools

Step 4 Run the following command to change the password for user **ommdba**:

mod_db_passwd ommdba

Step 5 Enter the old password of user **ommdba** and enter a new password twice.

NOTE

The default password of the OMS database administrator **ommdba** is randomly generated by the system. For details about how to obtain a random password, see ***Huawei Cloud Stack 8.3.1 Account List*** .

The password must meet the following complexity requirements:

- Contains 16 to 32 characters.
- Contains at least three types of the following: uppercase letters, lowercase letters, numbers, and special characters (~`!@#$%^&*()-+_=\|[{}];,<.>/?).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be the same as the last 20 historical passwords.

If the following information is displayed, the password is changed:

```
Congratulations, update [ommdba] password successfully.
```

----End

3.2.3.3.2 Changing the Password for the Data Access User of the OMS Database

Scenario

It is recommended that the administrator periodically change the password of the user accessing the OMS database to improve the system O&M security.

Impact on the System

The OMS service needs to be restarted for the new password to take effect. The service is unavailable during the restart.

Procedure

Step 1 On FusionInsight Manager, choose **System** > **OMS** > **gaussDB** > **Change Password**.

Step 2 Locate the row where user **omm** is located and click **Change Password** in the **Operation** column.

Step 3 In the displayed dialog box, enter the password of the current login user and click **OK**.

Step 4 Enter the old and new passwords as prompted.

NOTE

The default password of the OMS database user **omm** is randomly generated by the system. For details about how to obtain a random password, see ***Huawei Cloud Stack 8.3.1 Account List*** .

The password must meet the following complexity requirements:

- Contains 8 to 32 characters.
- Contains at least three types of the following: uppercase letters, lowercase letters, numbers, and special characters (~`!@#$%^&*()-+_=\|[{}];,<.>/?).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be the same as the last 20 historical passwords.

Step 5 Click **OK**. Wait until the system displays a message indicating that the operation is successful.

Step 6 Locate the row where user **omm** is located and click **Restart OMS Service** in the **Operation** column.

Step 7 In the displayed dialog box, enter the password of the current login user and click **OK**.

Step 8 In the displayed restart confirmation dialog box, click **OK** to restart the OMS service.

----End

3.2.3.3.3 Resetting the Component Database User Password

Scenario

Default passwords for components in the MRS cluster to connect to the DBService database are random. You are advised to periodically reset the passwords of component database users to improve system O&M security.

Impact on the System

To reset passwords, you need to stop and then restart services, during which services are unavailable.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Cluster** > **Services**.

Step 2 Click the name of the service whose database user password is to be reset, for example, **Kafka**, and click **Stop Service** on the **Dashboard** page.

In the displayed dialog box, enter the password of the current login user and click **OK**.

After confirming the impact of stopping the service, wait until the service is stopped.

Step 3 On the **Dashboard** page, choose **More** > **Reset Database Password**.

In the displayed dialog box, enter the password of the current login user and click **OK**.

Select "I have read the information and understand the impact", and click **OK**.

Step 4 After the password is reset, click **Start Service** on the **Dashboard** page.

Step 5 In the displayed dialog box, click **OK** and wait until the service is started.

----End

3.2.3.3.4 Resetting the Password for User omm in DBService

Scenario

The default password of the DBService database user **omm** in the MRS cluster is randomly generated. Periodically reset the password to improve system O&M security.

Impact on the System

Services need to be stopped and restarted and therefore become unavailable during this period.

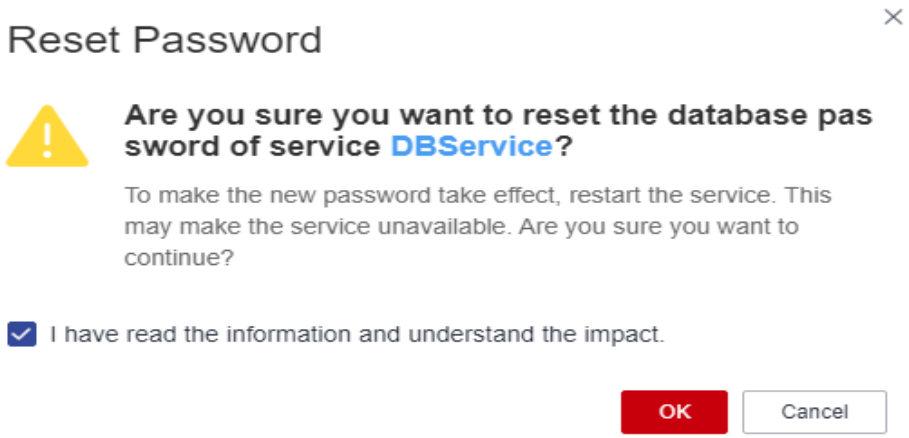
Procedure

Step 1 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **DBService**.

Step 2 On the **Dashboard** page, click **More** and select **Reset Database Password**.

In the displayed dialog box, enter the password of the current login user and click **OK**.

Select **I have read the information and understand the impact**, and click **OK**.



Step 3 After the password is reset, click **More** and select **Restart Service** on the **Dashboard** page.

In the displayed dialog box, enter the password of the current login user and click **OK**.

Confirm the impact of restarting the service, click **OK**, and wait until the service is started.

----End

3.2.3.3.5 Changing the Password for User compdbuser of the DBService Database

Scenario

It is recommended that the administrator periodically change the password of the OMS database administrator to improve the system O&M security.

Procedure

Step 1 Log in to FusionInsight Manager, choose **Cluster** > **Services** > **DBService**, click **Instance**, and view the IP address of the active DBService node.

Step 2 Log in to the active DBService node as user **root**.

NOTE

The password of user **compuserdb** cannot be changed on the standby DBService node. Change the password on the active management node only.

Step 3 Switch to the **$DBSERVER_HOME** directory and configure environment variables:

su - omm

cd $DBSERVER_HOME

source .dbservice_profile

Step 4 Run the following command to change the password of user **compdbuser** as user **omm** of the DBService database:

gsql -U omm -W omm*Password of user omm of the DBService database* **-d postgres -p 20051 -c "alter user compdbuser identified by '***New password***' valid until '***Expiration time***';"**

NOTE

- The default password of the DBService database user **omm** is randomly generated by the system. For details about how to obtain a random password, see ***Huawei Cloud Stack 8.3.1 Account List*** .
- The new password must meet the following complexity requirements:
  - The password contains 16 to 32 characters.
  - The password contains at least three types of the following: uppercase letters, lowercase letters, digits, and special characters ('~!@#$%^&*()-_=+\|[{}];:'",<.>/?).
  - The password cannot be the same as the username or the username spelled backwards.
  - The password cannot be the same as the last 20 historical passwords.
- The expiration time format is xxxx-xx-xx, for example, **2020-10-31**.

If the following information is displayed, the modification is successful:

```
ALTER ROLE
```

----End

3.2.3.4 Changing Passwords on the Management and Control Plane

For details about Huawei Cloud Stack account management policies and how to change passwords of various users, see "O&M Guide" > "Security Management" > "Account Management" in *Huawei Cloud Stack 8.3.1 Product Documentation* .

Changing Passwords Using the Account Management Function of ManageOne

If you need to maintain passwords during security maintenance, you can quickly change the passwords on ManageOne Maintenance Portal to improve O&M efficiency and enhance account password security. For details, see "Changing Account Passwords Using Accounts" in *Huawei Cloud Stack 8.3.1 Product Documentation* .

Changing Passwords for MySQL Accounts

NOTE

The following describes how to change the password for MRS management plane services after the password of the MRS-DB database user is changed.

Step 1 Log in to the **EICommon-Region-Master-01** VM as user **opsadmin** by referring to **Logging In to an MRS Management Node** and then switch to user **root**.

su - root

See ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator to obtain the default password.

Step 2 Run the following command to query the names of all MRS containers in the cluster:

kubectl get pods -n mrs -o wide

```
NAME                          READY   STATUS    RESTARTS   AGE    IP           NODE          NOMINATED
NODE   READINESS GATES
mrsapigw-6f56bc476d-c5cs9      1/1     Running   0          19h    10.16.0.69    10.69.26.187   <none>
<none>
mrsapigw-6f56bc476d-smqn8      1/1     Running   0          5d13h   10.16.0.20    10.69.26.197
<none>          <none>
mrsdeployer-5988d78867-2prk6   1/1     Running   0          43h    10.16.0.55    10.69.26.194   <none>
<none>
mrsdeployer-5988d78867-8lb5p   1/1     Running   0          43h    10.16.0.25    10.69.26.197   <none>
<none>
mrsdeployer-5988d78867-gwqb7   1/1     Running   0          43h    10.16.0.86    10.69.26.189
<none>          <none>
```

Run the following command to log in to any MRS-Deployer container node, for example, **mrsdeployer-5988d78867-2prk6**:

kubectl exec -it -n mrs *mrsdeployer-5988d78867-2prk6* **bash**

Step 3 Run the following commands in sequence to encrypt the plaintext password to be changed and record the encrypted password:

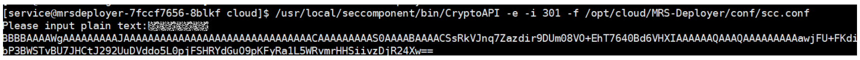
New deployment scenario:

/usr/local/seccomponent/bin/CryptoAPI -e -i 301 -f /opt/cloud/MRS-Deployer/conf/scc.conf

Upgrade to 8.3.1:

/usr/local/seccomponent/bin/CryptoAPI -e -i 0 -f /opt/cloud/MRS-Deployer/conf/scc.conf

The following figure shows how to encrypt a password.



Run the **exit** command to exit MRS-Deployer.

Step 4 Run the following command to log in to any MRS-Apigw container node, for example, **mrsapigw-6f56bc476d-c5cs9**:

kubectl exec -it -n mrs *mrsapigw-6f56bc476d-c5cs9* **bash**

Step 5 Run the following commands in sequence to encrypt the plaintext password to be changed and record the encrypted password:

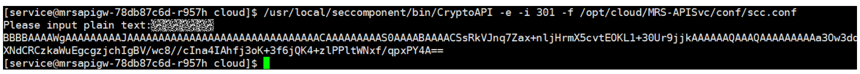
New deployment scenario:

/usr/local/seccomponent/bin/CryptoAPI -e -i 301 -f /opt/cloud/MRS-APISvc/conf/scc.conf

Upgrade to 8.3.1:

/usr/local/seccomponent/bin/CryptoAPI -e -i 0 -f /opt/cloud/MRS-APISvc/conf/scc.conf

The following figure shows how to encrypt a password.



Step 6 Open a browser, enter https://*CloudScopeLite domain name* in the address box, and press **Enter** to access CloudScope.

Log in to CloudScope as user **op_cdk_sso**. Contact the system administrator to obtain the login password.

Step 7 On the **Services** > **Change Mgmt** tab page, click **CloudAutoDeploy-CDK**.

Choose **Change Mgmt** > **Upgrade**, select **ei-dbs-region** for **Cluster** in the upper right corner, and set **Namespace** to **mrs**. Search for **mrsapigw-*XXX*** and **mrsdeployer-*XXX***, and click **Next**. On the **Upgrade Configuration** page, modify parameters by referring to the following table.

Table 3-3 Microservice parameter configuration items

Parameter	Value	Example	Description
CC_DB_PASSWD	Encrypted password of MRS-Deployer.	See the encrypted password returned in **Step 3**.	Corresponds to the parameters on the **Upgrade Configuration** page of all mrsdeployer-*XXX* instances.
CC_DB_WCC_PASSWD	Encrypted password of MRS-Apigw.	See the encrypted password returned in **Step 5**.	Corresponds to the parameters on the **Upgrade Configuration** page of all mrsapigw-*XXX* instances.
POD_RESTART	Used to forcibly restart the POD.	2	Corresponds to the parameters on the **Upgrade Configuration** page of the mrsdeployer-*XXX* and mrsapigw-*XXX* instances. This parameter is used to forcibly restart the POD. Each time other parameters are modified, this parameter must be modified. You can set this parameter to any value.

On the **Upgrade Configuration** page, after the parameters are modified, click **Next**, and click **Upgrade**.

NOTE

If MRS is deployed across AZs on the management plane, modify the parameters of the **ei-dbs-region-dr** cluster in the same way after modifying the parameters of the **ei-dbs-region** cluster.

----End

3.2.4 Security Policies

For details about Huawei Cloud Stack security policies and other settings, see "Maintenance Guide" > "O&M Guide" > "Security Management" in *Huawei Cloud Stack 8.3.1 Product Documentation* .

3.3 Certificate Management

3.3.1 Certificate List

For details about the certificates of the MRS management and control plane and MRS clusters, see *Huawei Cloud Stack 8.3.1 Certificate List* .

3.3.2 Certificate Replacement

3.3.2.1 Replacing the CA Certificate for an MRS Cluster

Scenario

The MRS CA certificate is used to encrypt data during the communication between the client and server of a component to ensure secure communication. You can replace the CA certificate on FusionInsight Manager to ensure product security. This operation is applicable to the following scenarios:

- After the cluster is installed for the first time, import an enterprise certificate.
- If the enterprise certificate has expired or security hardening is required, replace it with a new certificate.

After the CA certificate is replaced, the certificates that are used by HDFS, Yarn, MapReduce, HBase, Loader, Hue, Flink, FTP-Server, Oozie, Hive, Guardian, Tomcat, CAS, HTTPD, and LDAP in MRS will be automatically updated.

The certificate file and key file can be applied for from the enterprise certificate center or generated by the cluster user.

NOTE

- Only CA certificates that can be issued and in **X.509** format can be imported in FusionInsight.
- FusionInsight requires that the OS encoding format be **en_US.UTF-8** or **POSIX**. Otherwise, the certificate function will be abnormal.
- If an isolated faulty node exists in the current cluster, the CA certificate of the node will not be replaced. After the node is de-isolated, you need to reinstall the services running on the node to ensure that the node and the cluster use the same CA certificate.

Impact on the System

During the replacement, the MRS system needs to be restarted and cannot be accessed or provide services.

Prerequisites

- You have obtained the files to be imported to the MRS cluster, including the CA certificate file (***.crt**), key file (***.key**), and file that saves the key file password (**password.property**). The certificate name and key name support letters and digits.
- You have prepared a password for accessing the key file.

  To avoid potential security risks, the password must meet the following complexity requirements:

  – Must contain at least 8 characters.

  – Must contain at least four types of the following characters: uppercase letters, lowercase letters, digits, and special characters (~`!?,.;-_'(){}[]/<>@#$%^&*+|\=).
- When applying for a certificate from the certificate center, provide the password for accessing the key file and apply for the certificate files in CRT, CER, CERT, and PEM formats and the key files in KEY and PEM formats. The applied certificate must have the issuing function.

Procedure

Step 1 Log in to any management node in the cluster as user **omm**.

Step 2 Select a method for generating certificate files and key files.

- If the certificate is generated by the certificate center, save the certificate file and key file to the **omm** user directory on the management node.

  NOTE

  If the obtained certificate file is not in the **.crt** format and the key file is not in the **.key** format, run the following commands to change the file formats:

  mv *Certificate name.Certificate formatCertificate name***.crt**

  mv *Key name.Key format Key name***.key**

  For example, run the following commands to name the certificate file **ca.crt** and name the key file **ca.key**:

  mv server.cer ca.crt

  mv server_key.pem ca.key

- If the certificate is generated by the cluster user, run the following commands to generate the certificate file and key file in the **omm** user directory on the management node:

  NOTE

  To issue a root CA certificate with the KeyUsage field, modify the keyUsage field under [v3_ca] in the **openssl.cnf** file.

  - SLES system file path: **/etc/ssl/openssl.cnf**
  - CentOS and EulerOS system file path: **/etc/pki/tls/openssl.cnf**

  For example: keyUsage = cRLSign, keyCertSign

  a. Generate the key file.

  Run the following command to check whether the OpenSSL version is 1.1.1 or later:

  /usr/bin/openssl version

  - If yes, run the following command:

    openssl genrsa -out *Key name*.**key -aes256 3072**

  - If no, run the following command:

    openssl genrsa -out *Key name*.**key -aes256 3072 -sha256**

  If key file **ca.key** is generated and the OpenSSL version is earlier than 1.1.1, run the following command:

  openssl genrsa -out ca.key -aes256 3072 -sha256

  Enter the password twice as prompted, and press **Enter**.

  ```
  Enter pass phrase for ca.key:
  Verifying - Enter pass phrase for ca.key:
  ```

  b. Generate the certificate file.

  openssl req -new -x509 -days 1825 -key *Key name*.**key -out** *Certificate name*.**crt -subj "/C=cn/ST=guangdong/L=shenzhen/O=huawei/OU=huawei/CN=huawei" -sha256**

  For example, to generate the certificate file **ca.crt**, run the following command:

  openssl req -new -x509 -days 1825 -key ca.key -out ca.crt -subj "/C=cn/ST=guangdong/L=shenzhen/O=huawei/OU=huawei/CN=huawei" -sha256

  Enter the password for the key file as prompted, and press **Enter**.

  ```
  Enter pass phrase for ca.key:
  ```

Step 3 Run the following command in the **omm** user directory on the management node to save the password for accessing the key file.

sh ${BIGDATA_HOME}/om-server/om/sbin/genPwFile.sh

Enter the password twice as prompted, and press **Enter**. After being encrypted, the password is saved in **password.property**.

```
Please input key password:
Please Confirm password:
```


NOTE

- The **password.property** file generated on the node you have logged in is available only for the current cluster and cannot be used for other clusters. The file contains security information. Keep it secure and control the access permission.
- In active/standby DR scenarios, the **genPwFile.sh** script must be executed on both the active and DR cluster nodes, and the same password must be entered for the two clusters.

Step 4 Compress the three files in the **.tar** format and save them to the local computer.

tar -cvf *Package name Certificate name* **.crt** *Key name* **.key password.property**

For example, **tar -cvf test.tar ca.crt ca.key password.property**

NOTE

In active/standby DR scenarios, run this command on each cluster node.

Step 5 Log in to FusionInsight Manager and choose **System** > **Certificate**.

Step 6 In the **Upload Certificate** area, click the file selection button. In the window for selecting files, select the obtained **.tar** certificate file packages and open them and click **Upload**. The system automatically imports the certificate.

Step 7 After the certificate is imported, the system prompts you to synchronize the cluster configuration and restart the web service for the new certificate to take effect. After you complete these operations, click **OK**.

Step 8 In the dialog box that is displayed, enter the password and click **OK** to automatically synchronize the cluster configuration and restart the web service.

Step 9 After the cluster is restarted, enter the URL for accessing FusionInsight Manager in the address box of the browser and check whether the FusionInsight Manager web UI can be successfully displayed.

NOTE

If the enterprise certificate has expired or security is enhanced, replace the local certificate after replacing the MRS certificate. For details, see **Installing Cluster Certificates**.

Step 10 In the upper right corner of **Homepage**, click **More** and select **Restart**. In the dialog box displayed, enter the password of the current login user and click **OK**.

NOTE

After the CA certificate is replaced, you need to restart the cluster offline to make the certificate take effect. Rolling restart is not supported.

Step 11 In the displayed restart confirmation dialog box, click **OK**.

----End

3.3.2.2 Replacing the HA Certificate for an MRS Cluster

Scenario

HA certificates are used to encrypt the communication between active/standby processes and high availability processes to ensure security. Replace the HA

certificates on active and standby management nodes on FusionInsight Manager to ensure product security. This operation is applicable to the following scenarios:

- After the cluster is installed for the first time, import an enterprise certificate.
- If the enterprise certificate has expired or security hardening is required, replace it with a new certificate.

NOTE

This section applies only to physical machine clusters. but is not applicable to scenarios where active and standby management nodes are not installed.

The certificate file and key file can be applied for from the enterprise certificate center or generated by the cluster user.

Impact on the System

FusionInsight Manager must be restarted during the replacement and cannot be accessed or provide services.

Prerequisites

- You have obtained the **root-ca.crt** root file and the **root-ca.pem** key file of the certificate to be replaced.
- You have prepared a password for accessing the key file.

  To avoid potential security risks, the password must meet the following complexity requirements:

  - Must contain at least 8 characters.
  - Must contain at least four types of the following characters: uppercase letters, lowercase letters, digits, and special characters (~`!?,.;-_'(){}[]/<>@#$%^&*+|\=).
- When applying for a certificate from the certificate center, provide the password for accessing the key file and apply for the certificate files in CRT, CER, CERT, and PEM formats and the key files in KEY and PEM formats. The applied certificate must have the issuing function.

Procedure

Step 1 Log in to the active management node as user **omm** using the IP address of the active management node.

Step 2 Select a method for generating certificate files and key files.

- If the certificate is generated by the certificate center, save the certificate file and key file to the **${OMS_RUN_PATH}/workspace0/ha/local/cert** directory on the active and standby management nodes.

NOTE

If the obtained certificate file is not in the **.crt** format and the key file is not in the **.pem** format, run the following commands to change the file formats:

mv *Certificate name.Certificate format***root-ca.crt**

mv *Key name.Key format* **root-ca.pem**

For example, run the following commands to name the certificate file **root-ca.crt** and the key file **root-ca.pem**:

mv server.cer root-ca.crt

mv server_key.key root-ca.pem

- If the certificate is generated by the cluster user, run the following command to generate **root-ca.crt** and **root-ca.pem** in the **${OMS_RUN_PATH}/workspace0/ha/local/cert** directory:

  sh ${OMS_RUN_PATH}/workspace/ha/module/hacom/script/gen-cert.sh --root-ca --country=CN --state=*state* **--city=***city* **--company=***company* **--organize=***organize* **--common-name=***commonname* **--email=***Cluster user email address*

  NOTE

  The validity period of the generated certificate file is 5 years. When the system certificate file is about to expire, the system generates the "ALM-12055 Certificate File Is About to Expire" alarm.

  For example, run the following command:

  sh ${OMS_RUN_PATH}/workspace/ha/module/hacom/script/gen-cert.sh --root-ca --country=CN --state=guangdong --city=shenzhen --company=huawei --organize=IT --common-name=HADOOP.COM --email=abc@*xxx*.com

  Enter the password as prompted and press **Enter**.

  ```
  Enter pass phrase for /opt/huawei/Bigdata/om-server/OMS/workspace/ha/local/cert/root-ca.pem:
  ```

  The command is executed if the following information is displayed:

  ```
  Generate root-ca pair success.
  ```

Step 3 On the active management node, run the following command as user **omm** to copy **root-ca.crt** and **root-ca.pem** to the **${BIGDATA_HOME}/om-server/om/security/certHA** directory:

cp -arp ${OMS_RUN_PATH}/workspace0/ha/local/cert/root-ca.* ${BIGDATA_HOME}/om-server/om/security/certHA

Step 4 Copy **root-ca.crt** and **root-ca.pem** generated on the active management node to the **${BIGDATA_HOME}/om-server/om/security/certHA** directory on the standby management node as user **omm**.

scp ${OMS_RUN_PATH}/workspace0/ha/local/cert/root-ca.* omm@***IP address of the standby management node***:${BIGDATA_HOME}/om-server/om/security/certHA

Step 5 Run the following command to generate an HA certificate and perform the automatic replacement:

sh ${BIGDATA_HOME}/om-server/om/sbin/replacehaSSLCert.sh

Enter the password as prompted and press **Enter**.


```
Please input ha ssl cert password:
```

The DBService HA certificate is replaced successfully if the following information is displayed:

```
[INFO] Succeed to replace ha ssl cert.
```

NOTE

If the user wants to update the package for encrypting the HA password, add the **-u** parameter.

Step 6 Run the following command to restart the OMS:

sh ${BIGDATA_HOME}/om-server/om/sbin/restart-oms.sh

The following information is displayed:

```
start HA successfully.
```

Step 7 Log in to the standby management node as user **omm** using the IP address of the standby management node.

Run **sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh** to check whether **HAAllResOK** of the management node is **Normal** and whether FusionInsight Manager can be logged in to again. If yes, the operation is successful.

----End

3.3.2.3 Replacing OMS Gaussdb Certificates

Scenario

The Gaussdb certificate is used to enable SSL encrypted transmission between the active and standby OMS databases. This section provides guidance on how to replace Gaussdb certificate files on the active and standby OMS nodes on FusionInsight Manager in the following scenarios:

- If the enterprise certificate has expired or security hardening is required, replace it with a new certificate.

NOTE

This section applies only to physical machine clusters. But it does not apply to scenarios where active and standby management nodes are not installed.

The certificate file and key file can be applied for from the enterprise certificate administrator or generated by the system administrator.

Impact on the System

OMS Gaussdb needs to be restarted during the replacement and cannot be accessed or provide services.

Prerequisites

- You have obtained the certificate files.
- You have prepared a password for accessing the key file.

To avoid potential security risks, the password must meet the following complexity requirements:

- Must contain at least 8 characters.
- Contains at least three types of the following: uppercase letters, lowercase letters, digits, and special characters (~!@#$%^&*-_=+|[];,<.>/?).

- When applying for a certificate from the certificate administrator, provide the password for accessing the key file and apply for the certificate files in CRT, CER, CERT, and PEM formats and the key files in KEY and PEM formats. The applied certificate must have the issuing function.

Procedure

Step 1 Use PuTTY to log in to the active OMS node as user **root**.

Step 2 Select a method for generating certificate files and key files.

- If the certificate file is generated by the certificate administrator, save the certificate file and key file to the **/home/ommdba/** directory on the active and standby management nodes.

  NOTE

  If the obtained certificate file is not in the **.crt** format and the key file is not in the **.pem** format, run the following commands to change the file formats:

  mv *Certificate name.Certificate format***root-ca.crt**

  mv *Key name.Key format* **root-ca.pem**

  For example, run the following commands to name the certificate file **root-ca.crt** and the key file **root-ca.pem**:

  mv server.cer root-ca.crt

  mv server_key.key root-ca.pem

- If the certificate file is generated by the system administrator, run the following command to generate the **root-ca.crt** and **root-ca.pem** files:

  sh ${OMS_RUN_PATH}/workspace/ha/module/hacom/script/gen-cert.sh --root-ca --country=CN --state=*state* **--city=***city* **--company=***company* **--organize=***organize* **--common-name=***commonname* **--email=***Administrator email address* **--root-ca-file=/home/ommdba/root-ca.crt --root-key-file=/home/ommdba/root-ca.pem**

  NOTE

  The validity period of the generated certificate file is 5 years. When the system certificate file is about to expire, the system generates the "ALM-12055 Certificate File Is About to Expire" alarm.

  For example, run the following command:

  sh ${OMS_RUN_PATH}/workspace/ha/module/hacom/script/gen-cert.sh --root-ca --country=CN --state=guangdong --city=shenzhen --company=huawei --organize=IT --common-name=HADOOP.COM --email=abc@*xxx***.com --root-ca-file=/home/ommdba/root-ca.crt --root-key-file=/home/ommdba/root-ca.pem**

  Enter the password as prompted and press **Enter**.

  ```
  Enter pass phrase for /home/ommdba/root-ca.pem:
  ```

  The command is executed if the following information is displayed:


```
Generate root-ca pair success.
```

Step 3 Issue the OMS GaussDB certificate and run the following commands to generate the **server.crt** and **serve.pem** files:

sh ${OMS_RUN_PATH}/workspace/ha/module/hacom/script/gen-cert.sh --server-ca --country=CN --state=*state* **--city=***city* **--company=***company* **--organize=***organize* **--common-name=***commonname* **--email=***Administrator email address* **--root-ca-file=/home/ommdba/root-ca.crt --root-key-file=/home/ommdba/root-ca.pem --server-ca-file=/home/ommdba/server.crt --server-key-file=/home/ommdba/server.pem**

For example, run the following command:

sh ${OMS_RUN_PATH}/workspace/ha/module/hacom/script/gen-cert.sh --server-ca --country=CN --state=guangdong --city=shenzhen --company=huawei --organize=IT --common-name=HADOOP.COM --email=abc@*xxx*.com --root-ca-file=/home/ommdba/root-ca.crt --root-key-file=/home/ommdba/root-ca.pem --server-ca-file=/home/ommdba/server.crt --server-key-file=/home/ommdba/server.pem

Enter the password as prompted and press **Enter**.

```
Enter pass phrase for /home/ommdba/server.pem:
Enter pass phrase for /home/ommdba/root-ca.pem:
```

The command is executed if the following information is displayed:

```
Generate server-ca pair success.
```

Step 4 Run the following commands to view the certificate file and change the owner to user **ommdba**:

cd /home/ommdba/

chown ommdba: root-ca.crt root-ca.pem server.pem server.crt

Step 5 Replace the certificate file as user **ommdba**.

1. Run the following command to switch to user **ommdba**:

   su - ommdba

2. Go to the **${GAUSSHOME}/../** directory and run the following commands to replace the certificate file:

   cd ${GAUSSHOME}/../

   sh replace_database_ssl.sh *<Certificate file> <Key file> <Root certificate file>*

   Example:

   sh replace_database_ssl.sh /home/ommdba/server.crt /home/ommdba/server.pem /home/ommdba/root-ca.crt

   Enter the password as prompted and press **Enter**.

   ```
   Enter encryption password of the key file:
   Enter password again:
   ```

   The certificate file is replaced if the following information is displayed:

   ```
   Replace gaussdb SSL Certificates files successfully.
   ```

   NOTE

   The file specified in the parameter must contain an absolute path.

Step 6 Log in to the standby OMS node and perform **Step 2** to **Step 5**.

----End

3.3.2.4 Replacing DBService HA Certificates

Scenario

The DBService HA certificate is used to encrypt the data between active/standby processes and HA processes to ensure secure communications. This section provides guidance on how to replace DBService HA certificate files on the active and standby DBServer nodes in the following scenarios:

- After the cluster is installed for the first time, import an enterprise certificate.
- If the enterprise certificate has expired or security hardening is required, replace it with a new certificate.

NOTE

This section applies only to physical machine clusters. but is not applicable to scenarios where active and standby management nodes are not installed.

The certificate file and key file can be applied for from the enterprise certificate administrator or generated by the system administrator.

Impact on the System

DBService needs to be restarted during the replacement and cannot be accessed or provide services at that time.

Prerequisites

- You have obtained the **root-ca.crt** root file and the **root-ca.pem** key file of the certificate to be replaced.
- You have prepared a password for accessing the key file.

  To avoid potential security risks, the password must meet the following complexity requirements:

  - Must contain at least 8 characters.
  - Must contain at least four types of the following characters: uppercase letters, lowercase letters, digits, and special characters (~`!?,.;-_'(){}[]/<>@#$%^&*+|\=).
- When applying for a certificate from the certificate administrator, provide the password for accessing the key file and apply for the certificate files in CRT, CER, CERT, and PEM formats and the key files in KEY and PEM formats. The applied certificate must have the issuing function.

Procedure

Step 1 Log in to Manager, choose **Cluster** > **Services** > **DBService**, and click the **Instance** tab. On the displayed page, view the IP addresses of the active and standby DBServer nodes.

Step 2 Log in to the active DBServer node using PuTTY as user **omm**.

Step 3 Select the certificate file and key file generation mode.

- If the certificate is generated by the certificate administrator, save the certificate file and key file to the **${DBSERVER_HOME}/ha/local/cert** directory on the active and standby management nodes.

  NOTE

  If the obtained certificate file is not in the **.crt** format and the key file is not in the **.pem** format, run the following commands to change the file formats:

  mv *Certificate name.Certificate format* **root-ca.crt**

  mv *Key name.Key format* **root-ca.pem**

  For example, run the following command to name the certificate file **root-ca.crt** and name the key file **root-ca.pem**:

  mv server.cer root-ca.crt

  mv server_key.key root-ca.pem

- If the certificate is generated by the system administrator, run the following command to generate the **root-ca.crt** and **root-ca.pem** files in the **${DBSERVER_HOME}/ha/local/cert** directory:

  sh ${DBSERVER_HOME}/ha/module/hacom/script/gen-cert.sh --root-ca --country=CN --state=*state* **--city=***city* **--company=***company* **--organize=***organize* **--common-name=***commonname* **--email=***Administrator email address*

  NOTE

  The validity period of the generated certificate file is 5 years. When the system certificate file is about to expire, the system generates the alarm, ALM-12055 Certificate File Is About to Expire.

  For example, run the following command:

  sh ${DBSERVER_HOME}/ha/module/hacom/script/gen-cert.sh --root-ca --country=CN --state=guangdong --city=shenzhen --company=huawei --organize=IT --common-name=HADOOP.COM --email=abc@*xxx*.com

  Enter the password as prompted, and press **Enter**.

  ```
  Enter pass phrase for /opt/huawei/Bigdata/FusionInsight_BASE_xxx/install/FusionInsight-dbservice-2.7.0/ha/local/cert/root-ca.pem:
  ```

  The command is run successfully if the following information is displayed:

  ```
  Generate root-ca pair success.
  ```

Step 4 On the active node, run the following command as user **omm** to copy **root-ca.crt** and **root-ca.pem** to the **${DBSERVER_HOME}/security** directory:

cp -arp ${DBSERVER_HOME}/ha/local/cert/root-ca.* ${DBSERVER_HOME}/security

Step 5 Copy **root-ca.crt** and **root-ca.pem** generated on the active DBServer node to the **${DBSERVER_HOME}/security** directory on the standby DBServer node as user **omm**.

scp ${DBSERVER_HOME}/ha/local/cert/root-ca.* omm@***IP address of the standby DBServer node***:${DBSERVER_HOME}/security

Step 6 Run the following command to generate an HA certificate and perform the automatic replacement:

sh ${DBSERVER_HOME}/sbin/replacehaSSLCert.sh

Enter the password as prompted and press **Enter**.

```
Please input ha ssl cert password:
```

The DBService HA certificate is replaced successfully if the following information is displayed:

```
[INFO] Succeed to replace ha ssl cert.
```

NOTE

If the user wants to update the package for encrypting the HA password, add the **-u** parameter.

Step 7 Run the following command to restart HA:

sh ${DBSERVER_HOME}/ha/module/hacom/script/stop_ha.sh

sh ${DBSERVER_HOME}/ha/module/hacom/script/start_ha.sh

Step 8 Log in to the standby DBServer node using its IP address as user **omm** and repeat **Step 6** to **Step 7**.

----End

3.3.2.5 Replacing DBService Gaussdb Certificates

Scenario

The DBService Gaussdb certificate is used to enable SSL encrypted transmission between the active and standby DBService databases. This section provides guidance on how to replace Gaussdb certificate files on the active and standby DBService nodes in the following scenarios:

If the certificate has expired or security hardening is required, replace it with a new certificate.

NOTE

This section applies only to physical machine clusters. but is not applicable to scenarios where active and standby management nodes are not installed.

Impact on the System

DBService needs to be restarted during the replacement and cannot be accessed or provide services at that time.

Prerequisites

- The cluster has been installed.
- SSL encrypted transmission between the active and standby DBservice databases has been enabled.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **DBService**. On the **Dashboard** tab page that is displayed, choose **More** > **Synchronize Configuration** in the upper right corner. In the displayed **Synchronize Configuration** dialog box, click **OK** to synchronize configurations and re-issue the certificates.

Step 2 After the configurations are synchronized, choose **More** > **Restart Service**. In the **Verify Identity** dialog box displayed, enter the password of user **admin** and click **OK**. In the **Restart Service** dialog box that is displayed, select **Restart upper-layer services** and click **OK**.

Step 3 Check the certificate.

1. Log in to FusionInsight Manager, choose **Cluster** > **Services** > **DBService**, click the **Instance** tab, and view the IP address of the DBServer node.
2. Log in to the node in **Step 3.1** as user **root**, run the **su - omm** command to switch to user **omm**. Then run the **cd $DBSERVER_HOME/security** command to go to the **$DBSERVER_HOME/security** directory to view the **db-server.crt** and **db-server.pem** certificate files.

----End

3.3.2.6 Replacing FlinkServer HA Certificates

Scenario

The FlinkServer HA certificate is used to encrypt the data between active/standby processes and HA processes to ensure secure communications. Replace the FlinkServer HA certificates on the active and standby nodes to ensure product security. This operation applies to the following scenarios:

- After the cluster is installed for the first time, import an enterprise certificate.
- If the enterprise certificate has expired or security hardening is required, replace it with a new certificate.

NOTE

This section applies only to physical machine clusters. But it does not apply to scenarios where active and standby management nodes are not installed.

The certificate file and key file can be applied for from the enterprise certificate administrator or generated by the system administrator.

Impact on the System

The Flink service is inaccessible because it needs to be restarted during the replacement.

Prerequisites

- You have obtained the HA root certificate file **root-ca.crt** and key file **root-ca.pem** to be replaced.

- You have prepared a password for accessing the key file.

  To avoid potential security risks, the password must meet the following complexity requirements:

  – Contains at least 8 characters.

  – Contains at least four types of the following: uppercase letters, lowercase letters, numbers, and special characters (~`!?,.;-_'(){}[]/<>@#$%^&*+|\=).

- When applying for certificates from the certificate administrator, you have provided the password for accessing the key file and applied for the certificate files in CRT, CER, CERT, and PEM formats and the key files in KEY and PEM formats. The requested certificates must have the issuing function.

Procedure

Step 1 Log in to Manager, choose **Cluster** > **Services** > **Flink**, and click **Instance**. On the tab page that is displayed, view the IP addresses of the active and standby FlinkServer nodes.

Step 2 Log in to the active FlinkServer node as user **omm**.

Step 3 Select the certificate and key file generation mode.

- If the certificate file and key file are generated by the certificate administrator, save the files to the **${BIGDATA_HOME}/FusionInsight_Flink_*/install/FusionInsight-Flink-*/ha/local/cert** directory on the active management node.

  NOTE

  If the obtained certificate file is not in the **.crt** format or the key file is not in the **.pem** format, run either of the following commands accordingly to correct the format:

  mv *Certificate file name.Certificate file format* **root-ca.crt**

  mv *Key file name.Key file format* **root-ca.pem**

  For example, run the following commands to name the certificate file **root-ca.crt** and the key file **root-ca.pem**:

  mv server.cer root-ca.crt

  mv server_key.key root-ca.pem

- If the certificate file and key file are generated by the system administrator, run the following command to generate the **root-ca.crt** and **root-ca.pem** files in the **${BIGDATA_HOME}/FusionInsight_Flink_*/install/FusionInsight-Flink-*/ha/local/cert** directory on the active management node:

  sh ${BIGDATA_HOME}/FusionInsight_Flink_*/install/FusionInsight-Flink-*/ha/module/hacom/script/gen-cert.sh --root-ca --country=CN --state=*state* **--city=***city* **--company=***company* **--organize=***organize* **--common-name=***commonname* **--email=***Administrator email address*

  NOTE

  The validity period of the generated certificate file is 5 years. When the system certificate file is about to expire, the system generates the "ALM-45654 Flink HA Certificate File Is About to Expire" alarm.

  For example, run the following command:

  sh ${BIGDATA_HOME}/FusionInsight_Flink_*/install/FusionInsight-Flink-*/ha/module/hacom/script/gen-cert.sh --root-ca --country=CN --

state=guangdong --city=shenzhen --company=huawei --organize=IT --common-name=HADOOP.COM --email=abc@*xxx*.com

Enter the password as prompted and press **Enter**.

```
Enter pass phrase for /opt/huawei/Bigdata/FusionInsight_Flink_xxx/install/FusionInsight-Flink-x.x.x/ha/local/cert/root-ca.pem:
```

The command is successfully executed if the following information is displayed:

```
Generate root-ca pair success.
```

Step 4 Run the following command to generate an HA certificate and perform the automatic replacement:

sh ${BIGDATA_HOME}/FusionInsight_Flink_*/install/FusionInsight-Flink-*/flink/sbin/proceed_ha_ssl_cert.sh

The FlinkServer HA certificate is replaced successfully if the following information is displayed:

```
[INFO] Succeed to replace ha ssl cert.
```

Step 5 Run the following command to restart HA:

sh ${BIGDATA_HOME}/FusionInsight_Flink_*/install/FusionInsight-Flink-*/ha/module/hacom/script/stop_ha.sh

sh ${BIGDATA_HOME}/FusionInsight_Flink_*/install/FusionInsight-Flink-*/ha/module/hacom/script/start_ha.sh

Step 6 Log in to the standby FlinkServer node using its IP address as user **omm** and repeat **Step 4** to **Step 5**.

Step 7 Restart the Flink service.

Log in to FusionInsight Manager and choose **Cluster** > **Services** > **Flink**. On the displayed page, click **More** and select **Restart Service**. In the dialog box that is displayed, enter the administrator password, confirm the restart, and click **OK** to restart Flink.

----End

3.3.2.7 Replacing MOTService HA Certificates

Scenario

The MOTService HA certificate is used to encrypt the data between MOTService's active/standby processes and HA processes to ensure secure communications. Replace the MOTService HA certificates on the active and standby nodes to ensure product security. This operation applies to the following scenarios:

If the MOTService HA certificate has expired or security hardening is required, replace it with a new certificate.

NOTE

This section applies only to physical machine clusters. But it does not apply to scenarios where active and standby management nodes are not installed.

The certificate file and key file can be applied for from the enterprise certificate administrator or generated by the system administrator.

Impact on the System

The system is inaccessible and the MOTService service becomes unavailable because MOTService needs to be restarted during the replacement.

Prerequisites

- You have obtained the HA root certificate file **root-ca.crt** and key file **root-ca.pem** to be replaced.
- You have prepared a password for accessing the key file.

  To avoid potential security risks, the password must meet the following complexity requirements:

  – Contains at least 8 characters.

  – Contains at least four types of the following: uppercase letters, lowercase letters, numbers, and special characters (~`!?,.;-_'(){}[]/<>@#$%^&*+|\=).
- When applying for certificates from the certificate administrator, you have provided the password for accessing the key file and applied for the certificate files in CRT, CER, CERT, and PEM formats and the key files in KEY and PEM formats. The requested certificates must have the issuing function.

Procedure

Step 1 Log in to Manager, choose **Cluster** > **Services** > **MOTService**, and click **Instance**. On the tab page that is displayed, view the IP addresses of the active and standby MOTService nodes.

Step 2 Log in to the active MOTService node as user **omm**.

Step 3 Select the certificate and key file generation mode.

- If the certificate file and key file are generated by the certificate administrator, save the files to the **${MOTSERVER_HOME}/ha/local/cert** directory on the active and standby management nodes.

  NOTE

  If the obtained certificate file is not in the **.crt** format or the key file is not in the **.pem** format, run either of the following commands accordingly to correct the format:

  mv *Certificate file name.Certificate file format* **root-ca.crt**

  mv *Key file name.Key file format* **root-ca.pem**

  For example, run the following commands to name the certificate file **root-ca.crt** and the key file **root-ca.pem**:

  mv server.cer root-ca.crt

  mv server_key.key root-ca.pem
- If the certificate file and key file are generated by the system administrator, run the following command to generate the **root-ca.crt** and **root-ca.pem** files in the **${MOTSERVER_HOME}/ha/local/cert** directory on the active management node:

  sh ${MOTSERVER_HOME}/ha/module/hacom/script/gen-cert.sh --root-ca --country=CN --state=*state* **--city=***city* **--company=***company* **--**

organize=*organize* **--common-name=***commonname* **--email=***Administrator email address*

NOTE

The validity period of the generated certificate file is 5 years. When the system certificate file is about to expire, the system generates the "ALM-46010 MOTService Certificate File Is About to Expire" alarm.

For example, run the following command:

sh ${MOTSERVER_HOME}/ha/module/hacom/script/gen-cert.sh --root-ca --country=CN --state=guangdong --city=shenzhen --company=huawei --organize=IT --common-name=HADOOP.COM --email=abc@*xxx*.com

Enter the password as prompted and press **Enter**.

```
Enter pass phrase for /opt/huawei/Bigdata/FusionInsight_FARMER_RTD_8.*/install/FusionInsight-MOTService-*/ha/local/cert/root-ca.pem:
```

The command is successfully executed if the following information is displayed:

```
Generate root-ca pair success.
```

Step 4 On the active node, run the following command as user **omm** to copy **root-ca.crt** and **root-ca.pem** to the **${MOTSERVER_HOME}/security** directory:

cp -arp ${MOTSERVER_HOME}/ha/local/cert/root-ca.* ${MOTSERVER_HOME}/security

Step 5 Copy **root-ca.crt** and **root-ca.pem** generated on the active MOTService node to the **${MOTSERVER_HOME}/security** directory on the standby MOTService node as user **omm**.

scp ${MOTSERVER_HOME}/ha/local/cert/root-ca.* omm@***IP address of the standby MOTService node***:${MOTSERVER_HOME}/security

Step 6 Run the following command to load environment variables of MOTService:

source $MOTSERVER_HOME/.motservice_profile

Step 7 Run the following command to generate an HA certificate and perform the automatic replacement:

sh ${MOTSERVER_HOME}/sbin/proceed_ha_ssl_cert.sh ${MOTSERVER_HOME} ${MOTSERVER_IP}

Step 8 Run the following command to restart HA:

sh ${MOTSERVER_HOME}/ha/module/hacom/script/stop_ha.sh

sh ${MOTSERVER_HOME}/ha/module/hacom/script/start_ha.sh

Step 9 Log in to the standby MOTService node as user **omm** and repeat **Step 7** to **Step 8**.

Step 10 Restart the MOTService service.

Log in to FusionInsight Manager and choose **Cluster** > **Services** > **MOTService**. On the displayed page, click **More** and select **Restart Service**. In the dialog box that is displayed, enter the administrator password, confirm the restart, and click **OK** to restart MOTService.

----End

3.3.2.8 Replacing MOTService Certificates

Scenario

The MOTService certificate is used to enable SSL encrypted transmission between the active and standby MOTService databases. Replace the MOTService certificates on the active and standby nodes to ensure product security. This operation applies to the following scenarios:

If the certificate has expired or security hardening is required, replace it with a new certificate.

NOTE

This section applies only to physical machine clusters. But it does not apply to scenarios where active and standby management nodes are not installed.

Impact on the System

The system is inaccessible and the MOTService service becomes unavailable because MOTService needs to be restarted during the replacement.

Prerequisites

- The cluster has been installed.
- SSL encrypted transmission between the active and standby MOTService databases has been enabled.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **MOTService**. On the **Dashboard** tab page that is displayed, click **More** and select **Synchronize Configuration**. In the displayed **Synchronize Configuration** dialog box, click **OK** to synchronize configurations and re-issue the certificates.

Step 2 After the configurations are synchronized, click **More** and select **Restart Service**. In the **Verify Identity** dialog box displayed, enter the password of user **admin** and click **OK**. In the **Restart Service** dialog box that is displayed, select **Restart upper-layer services** and click **OK**.

Step 3 Check the certificate.

1. Log in to FusionInsight Manager, choose **Cluster** > **Services** > **MOTService**, click **Instance**, and view the IP address of the active node where MOTService is.
2. Log in to the node in **Step 3.1** as user **omm** and run the following command to load environment variables:

   source ${BIGDATA_HOME}/FusionInsight_FARMER_RTD_8.*/install/FusionInsight-MOTService-*/.motservice_profile
3. Run the following commands to go to the **${MOTSERVICE_HOME}/security** directory and view the **server.crt** certificate in the directory:

   cd ${MOTSERVICE_HOME}/security

ll

----End

3.3.2.9 Installing Cluster Certificates

Scenario

A public key certificate (certificate for short) is a digital signature statement that binds a public key to the ID of an individual, device, or service that has the related private key.

When you attempt to log in to FusionInsight Manager using a browser, the FusionInsight Manager home page may not be displayed due to a certificate error. In this case, install a certificate.

NOTE

This section applies only to physical machine clusters.

Procedure

In the following procedure, Google Chrome is used as an example to describe how to install a certificate.

1. In the address box of Google Chrome, enter the address for accessing FusionInsight Manager, and press **Enter**. On the page displayed, press **F12** to access Chrome DevTools. In the displayed window, click the **Security** tab then **View certificate**.

   NOTE

   Operations vary depending on Google Chrome versions. This following uses Google Chrome 110 as an example.

2. In the dialog box that is displayed, click **Details** and choose **huawei** in the **Certificate Hierarchy** area.
3. Click **Export...**. Enter the certificate file name in **File name**, for example, **D:\abc**, and set **Save as type** to **Base64-encoded ASCII, single certificate(*.pem;*.crt)**. Then, click **Save**.
4. In the upper right corner of Google Chrome, click ⋮ and choose **Settings**.
5. Choose **Privacy and security** > **Security**.
6. On the security page, choose **Manage device certificates**. On the **Personal** tab, click **Import**, and click **Next**.
7. In the **Certificate Import Wizard** dialog box, click **Browse**, select the certificate stored on the local host, for example, **D:\abc.crt**, click **Open**, and click **Next**.
8. In the **Certificate Store** dialog box, select **Place all certificates in the following store**, and click **Browse**.
9. Select **Trusted Root Certification Authorities**, and click **OK** to return to the **Certificate Import Wizard** dialog box.
10. Click **Next**, and then click **Finish**.

11. In the **Security Warning** dialog box that is displayed, click **Yes**. In the **The import was successful** dialog box that is displayed, click **OK**.
12. In **Privacy and security**, click **Clear browsing data**. In the dialog box that is displayed, select the items to be cleared and click **Clear data**.
13. Restart the browser and access FusionInsight Manager again.

3.3.2.10 Replacing the Deployer Service Certificate (Manual)

Prerequisites

The new **mrs.jks** certificate and password are available.

Procedure

Step 1 Log in to the **EICommon-Region-Master-01** VM as user **opsadmin** by referring to **Logging In to an MRS Management Node** and then switch to user **root**.

su - root

See ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator to obtain the default user password.

Step 2 Run the following command to query the names of all mrsdeployer containers in the cluster and obtain the IP addresses in the **NODE** column:

kubectl get pods -n mrs -owide

```
NAME                          READY   STATUS    RESTARTS   AGE     IP           NODE           NOMINATED
NODE   READINESS GATES
mrsapigw-6f56bc476d-c5cs9     1/1     Running   0          19h     10.16.0.69   10.69.26.187   <none>
<none>
mrsapigw-6f56bc476d-smqn8     1/1     Running   0          5d13h   10.16.0.20   10.69.26.197
<none>           <none>
mrsdeployer-5988d78867-2prk6  1/1     Running   0          43h     10.16.0.55   10.69.26.194   <none>
<none>
mrsdeployer-5988d78867-8lb5p  1/1     Running   0          43h     10.16.0.25   10.69.26.197   <none>
<none>
mrsdeployer-5988d78867-gwqb7  1/1     Running   0          43h     10.16.0.86   10.69.26.189
<none>           <none>
```

Step 3 Log in to the IP address (obtained in **Step 2**) of each MRS-Deploy node (whose container name starts with **mrsdeployer** in **Step 2**).

See ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator to obtain the default user password.

Step 4 Upload the new certificate to any directory on each MRS-Deploy node.

Step 5 Log in to each MRS-Deploy node as user **opsadmin**.

See ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator to obtain the default password of user **opsadmin**.

Run the **su - root** command to switch to user **root**.

Step 6 Run the **docker ps |grep deploy** command to view the container ID.

```
[opsadmin@MRS-DMZ-Node-03 ~]$ su - root
Password:
Last login: Thu Mar 18 15:26:16 CST 2021 on pts/3
[root@MRS-DMZ-Node-03 ~]# docker ps |grep deployer
7b8f995a283f        6cc8166b7c0b                                                "sh first_start.sh"   23 hours ago        Up 23
 hours                       k8s_mrsdeployer_mrsdeployer-598886b574-bmws8_mrs_b992bde5-6419-4836-9a9f-0f15cbf9
d606_0
8630556db9b4        cce-pause:3.1                                               "/pause"              23 hours ago        Up 23
 hours                       k8s_POD_mrsdeployer-598886b574-bmws8_mrs_b992bde5-6419-4836-9a9f-0f15cbf9d606_0
```


Step 7 Run the following commands to replace the certificate in the **/opt/cloud/MRS-Deployer/conf/certificate/** directory:

docker cp mrs.jks *Container ID*: **/tmp**

docker exec -u root -it *Container ID* **bash**

cp /tmp/mrs.jks /opt/cloud/MRS-Deployer/conf/certificate/

cd /opt/cloud/MRS-Deployer/conf/certificate/

chown service:servicegroup mrs.jks

chmod 600 mrs.jks

su - service

Container ID is the ID queried in **Step 6**.

Step 8 Run the following command to encrypt the password:

/usr/local/seccomponent/bin/CryptoAPI -e -f /opt/cloud/MRS-Deployer/conf/scc.conf

Enter the new certificate password.

```
Please input plain text:
```

Step 9 Replace the value of **ssl.server.keystore.password** in **/opt/cloud/MRS-Deployer/conf/core.properties** with the encrypted password.

Step 10 Restart the Deployer service.

sh /opt/cloud/MRS-Deployer/bin/stop.sh

sh /opt/cloud/MRS-Deployer/bin/start.sh

NOTE

If the Deployer container is restarted, repeat **Step 1** to **Step 10** to replace the certificate.

----End

3.3.2.11 Replacing the API Service Certificate (Manual)

Prerequisites

The new **mrs.jks** certificate and password are available.

Procedure

Step 1 Log in to the **EICommon-Region-Master-01** VM as user **opsadmin** by referring to **Logging In to an MRS Management Node** and then switch to user **root**.

su - root

See ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator to obtain the default user password.

Step 2 Run the following command to query the names of all mrsdeployer containers in the cluster and obtain the IP addresses in the **NODE** column:

kubectl get pods -n mrs -owide

```
NAME                                READY   STATUS    RESTARTS   AGE     IP           NODE           NOMINATED
NODE   READINESS GATES
mrsapigw-6f56bc476d-c5cs9           1/1     Running   0          19h     10.16.0.69   24.69.26.187   <none>
<none>
mrsapigw-6f56bc476d-smqn8           1/1     Running   0          5d13h   10.16.0.20   24.69.26.197
<none>           <none>
mrsdeployer-5988d78867-2prk6        1/1     Running   0          43h     10.16.0.55   24.69.26.194   <none>
<none>
mrsdeployer-5988d78867-8lb5p        1/1     Running   0          43h     10.16.0.25   24.69.26.197   <none>
<none>
mrsdeployer-5988d78867-gwqb7        1/1     Running   0          43h     10.16.0.86   24.69.26.189
<none>           <none>
```

Step 3 Log in to the IP address (obtained in **Step 2**) of each MRS-Api node (whose container name starts with **mrsapi** in **Step 2**).

See ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator to obtain the default user password.

Step 4 Upload the new certificate to any directory on each MRS-Api node.

Step 5 Log in to each MRS-Api node as user **opsadmin**.

See ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator to obtain the default password of user **opsadmin**.

Run the **su - root** command to switch to user **root**.

Step 6 Run the **docker ps |grep api** command to view the container ID.

```
[root@MRS-DMZ-Node-03 ~]# docker ps |grep api
a8fca9baba2d        172.35.134.14:20202/op_svc_cdk/mrsapigw-x86_64   "sh first_start.sh"   24 hours ago
mrs_cccac22c-e24d-4572-b3df-7b2c4fea2af3_0
1fd4bc9cb9b4        cce-pause:3.1                                    "/pause"              24 hours ago
ccac22c-e24d-4572-b3df-7b2c4fea2af3_0
```

Step 7 Run the following commands to replace the certificate in the **/opt/cloud/MRS-APISvc/conf/certificate** directory:

docker cp mrs.jks *Container ID*: **/tmp**

docker exec -u root -it *Container ID* **bash**

cp /tmp/mrs.jks /opt/cloud/MRS-APISvc/conf/certificate/

cd /opt/cloud/MRS-APISvc/conf/certificate

chown service:servicegroup mrs.jks

chmod 600 mrs.jks

su - service

Container ID is the ID queried in **Step 6**.

Step 8 Go to the **/opt/cloud/MRS-APISvc/conf/** directory and run the following command to obtain the ciphertext:

/usr/local/seccomponent/bin/CryptoAPI -e -f /opt/cloud/MRS-APISvc/conf/scc.conf

Enter the new certificate password.

```
Please input plain text:
```


Step 9 Change the value of **server.ssl.key-store** in **/opt/cloud/MRS-APISvc/conf/application.properties** to the new certificate path and the value of **server.ssl.key-store-password** to the encrypted password.

```
server.ssl.key-store=/opt/cloud/MRS-APISvc/conf/certificate/mrs.jks
server.ssl.key-store-password=xxx
```

NOTE

- Set **server.ssl.key-store** based on site requirements. Replace **mrs.jks** with the obtained certificate.
- Set **server.ssl.key-store-password** to the ciphertext obtained in **Step 8**.

Step 10 Restart the MRS-APi service.

sh /opt/cloud/MRS-APISvc/bin/stop.sh

sh /opt/cloud/MRS-APISvc/bin/start.sh

NOTE

If the Api container is restarted, repeat **Step 1** to **Step 10** to replace the certificate.

----End

3.3.2.12 Replacing the Deployer Service Certificate (Automatic)

MRS can automatically replace the Deployer certificate through ManageOne. The certificate name of MRS Deployer on ManageOne is **mrs-deployer-cdk-certfile**. For how to replace the certificate, see "O&M Guide" > "Security Management" > "Certificate Management" > "Replacing Type B and Type C Certificates" > "Replacing a Single or Multiple Certificates at a Time on ManageOne" in *Huawei Cloud Stack 8.3.1 Product Documentation* .

3.3.2.13 Replacing the API Service Certificate (Automatic)

MRS can automatically replace the API certificate through ManageOne. The certificate name of MRS Deployer on ManageOne is **mrs-apisvc-cdk-certfile**. For how to replace the certificate, see "O&M Guide" > "Security Management" > "Certificate Management" > "Replacing Type B and Type C Certificates" > "Replacing a Single or Multiple Certificates at a Time on ManageOne" in *Huawei Cloud Stack 8.3.1 Product Documentation* .

3.4 Security Hardening

3.4.1 Hardening Policies

Hardening Tomcat

Tomcat is hardened as follows based on open-source software during FusionInsight Manager software installation and use:

- Upgrade Tomcat to the official stable version apache-tomcat-9.0.83.
- Permissions on the directories under applications are set to **500**, and the write permission on some directories is supported.

- The Tomcat installation package is automatically deleted after the system software is installed.
- The automatic deployment function is disabled for projects in application directories. Only the **web**, **cas**, and **client** projects are deployed.
- Some unused **http** methods are disabled, preventing attacks by using the **http** methods.
- The default shutdown port and command of the Tomcat server are changed to prevent hackers from shutting down the server and attacking servers and applications.
- To ensure security, the value of **maxHttpHeaderSize** is changed, which enables server administrators to control abnormal requests of clients.
- The Tomcat version description file is modified after Tomcat is installed.
- To prevent disclosure of Tomcat information, the Server attributes of Connector are modified so that attackers cannot obtain information about the server.
- Permissions on files and directories of Tomcat, such as the configuration files, executable files, log directories, and temporary folders, are under control.
- Session facade recycling is disabled to prevent request leakage.
- LegacyCookieProcessor is used as CookieProcessor to prevent the leakage of sensitive data in cookies.

Hardening LDAP

LDAP is hardened as follows after a cluster is installed:

- In the LDAP configuration file, the password of the administrator account is encrypted using SHA. After the OpenLDAP is upgraded to 2.4.39 or later, data is automatically synchronized between the active and standby LDAP nodes using the SASL External mechanism, which prevents disclosure of the password.
- The LDAP service in the cluster supports the SSLv3 protocol by default, which can be used safely. When the OpenLDAP is upgraded to 2.4.39 or later, the LDAP automatically uses TLS1.0 or later to prevent unknown security risks.

Hardening JDK

- If the client process uses the AES256 encryption algorithm, JDK security hardening is required. The operations are as follows:

  Obtain the Java Cryptography Extension (JCE) package whose version matches that of JDK. The JCE package contains **local_policy.jar** and **US_export_policy.jar**. Copy the JAR files to the following directory and replace the files in the directory.

  - Linux: *JDK installation directory*/**jre/lib/security**
  - Windows: *JDK installation directory***jre\lib\security**

  NOTE

  Access the Open JDK open-source community to obtain the JCE file.

- If the client process uses the SM4 encryption algorithm, the JAR package needs to be updated.

Obtain **SMS4JA.jar** in the *client installation directory*/**JDK/jdk/jre/lib/ext/** directory, and copy the JAR package to the following directory:

- Linux: *JDK installation directory*/**jre/lib/ext/**
- Windows: *JDK installation directory***jre\lib\ext**

3.4.2 Configuring a Trusted IP Address to Access LDAP

Scenario

By default, the LDAP service deployed in the OMS and cluster can be accessed by any IP address. To enable the LDAP service to be accessed by only trusted IP addresses, you can configure the INPUT policy in the iptables filtering list.

Impact on the System

After the configuration, the LDAP service cannot be accessed by IP addresses that are not configured. Before the expansion, the added IP addresses need to be configured as trusted IP addresses.

Prerequisites

- You have collected the management plane IP addresses and service plane IP addresses of all nodes in the cluster and all floating IP addresses.
- You have obtained the **root** user account for all nodes in the cluster.

Procedure

Configuring trusted IP addresses for the LDAP service on the OMS

Step 1 Confirm the management node IP address. For details, see **Logging In to the Management Node**.

Step 2 Log in to FusionInsight Manager. For details, see **Logging In to FusionInsight Manager**.

Step 3 Choose **System** > **OMS** and choose **oldap** > **Modify Configuration** to view the OMS LDAP port number, that is, the value of **LDAP Listening Port**. The default port number is **21750**.

Step 4 Log in to the active management node as user **root** using the IP address of the active management node.

Step 5 Run the following command to check the INPUT policy in the iptables filtering list:

iptables -L

For example, if no rule is configured, the INPUT policy is displayed as follows:

```
Chain INPUT (policy ACCEPT)
target     prot opt source               destination
```

Step 6 Run the following command to configure all IP addresses used by the cluster as trusted IP addresses. Each IP address needs to be added independently.

iptables -A INPUT -s *Trusted IP address* **-p tcp --dport** *Port number* **-j ACCEPT**

For example, to configure **10.0.0.1** as a trusted IP address and enable it to access port **21750**, you need to run the following command:

iptables -A INPUT -s 10.0.0.1 -p tcp --dport 21750 -j ACCEPT

Step 7 Run the following command to configure all IP addresses as untrusted IP addresses. The trusted IP addresses will not be affected by this rule.

iptables -A INPUT -p tcp --dport *Port number* **-j DROP**

For example, to disable all IP addresses to access port **21750**, run the following command:

iptables -A INPUT -p tcp --dport 21750 -j DROP

Step 8 Run the following command to view the modified INPUT policy in the iptables filtering list:

iptables -L

For example, after a trusted IP address is configured, the INPUT policy is displayed as follows:

```
Chain INPUT (policy ACCEPT)
target     prot opt source               destination
ACCEPT     tcp  --  10.0.0.1            anywhere            tcp dpt:21750
DROP       tcp  --  anywhere            anywhere            tcp dpt:21750
```

Step 9 Run the following command to view the rules and rule numbers in the iptables filtering list:

iptables -L -n --line-number

```
Chain INPUT (policy ACCEPT)
num target     prot opt source               destination
1   DROP       tcp  --  0.0.0.0/0           0.0.0.0/0           tcp dpt:21750
```

Step 10 Run the following command to delete the desired rule from the iptables filtering list based on site requirement:

iptables -D INPUT *Number of the rule to be deleted*

For example, to delete rule 1, run the following command:

iptables -D INPUT 1

Step 11 Log in to the standby management node as user **root** using the standby IP address. Repeat **Step 5** to **Step 10**.

Configuring trusted IP addresses for the LDAP service in the cluster

Step 12 Log in to FusionInsight Manager.

Step 13 Choose **Cluster** > **Services** > **LdapServer**. Click **Instance** and view the LDAP nodes.

Step 14 Go to the **Configurations** page, and view the LDAP port number of the cluster, that is, the value of **LDAP_SERVER_PORT**. The default value is **21780**.

Step 15 Log in to the LDAP node as user **root** using the LDAP service IP address.

Step 16 Run the following command to view the INPUT policy in the iptables filtering list:

iptables -L

For example, if no rule is configured, the INPUT policy is displayed as follows:

```
Chain INPUT (policy ACCEPT)
target     prot opt source               destination
```

Step 17 Run the following command to configure all IP addresses used by the cluster as trusted IP addresses. Each IP address needs to be added independently.

iptables -A INPUT -s *Trusted IP address* **-p tcp --dport** *Port number* **-j ACCEPT**

For example, to configure **10.0.0.1** as a trusted IP address and enable it to access port **21780**, you need to run the following command:

iptables -A INPUT -s 10.0.0.1 -p tcp --dport 21780 -j ACCEPT

Step 18 Run the following command to configure all IP addresses as untrusted IP addresses. The trusted IP addresses will not be affected by this rule.

iptables -A INPUT -p tcp --dport *Port number* **-j DROP**

For example, to disable all IP addresses to access port **21780**, run the following command:

iptables -A INPUT -p tcp --dport 21780 -j DROP

Step 19 Run the following command to view the modified INPUT policy in the iptables filtering list:

iptables -L

For example, after a trusted IP address is configured, the INPUT policy is displayed as follows:

```
Chain INPUT (policy ACCEPT)
target     prot opt source               destination
ACCEPT     tcp  --  10.0.0.1             anywhere            tcp dpt:21780
DROP       tcp  --  anywhere             anywhere            tcp dpt:21780
```

Step 20 Run the following command to view the rules and rule numbers in the iptables filtering list:

iptables -L -n --line-number

```
Chain INPUT (policy ACCEPT)
num target     prot opt source               destination
1   DROP       tcp  --  0.0.0.0/0            0.0.0.0/0           tcp dpt:21780
```

Step 21 Run the following command to delete the desired rule from the iptables filtering list based on site requirement:

iptables -D INPUT *Number of the rule to be deleted*

For example, to delete rule 1, run the following command:

iptables -D INPUT 1

Step 22 Log in to the LDAP node as user **root** using the IP address of another LDAP service, and repeat **Step 16** to **Step 21**.

----End

3.4.3 Configuring NTP Security Authentication

Scenario

System allows a third-party NTP server to serve as an external clock source when Manager is installed to synchronize the time on the active management node with that on the external clock source. If NTP security authentication is not configured in a cluster, the time synchronization function is normal, but there may be security risks. To avoid the security risks, it is advised to use the external clock source with the enabled authentication function and to configure NTP security authentication.

By default, the active management node serves as an NTP server in the cluster. The time on the other nodes in the cluster is automatically synchronized with that on the active management node. NTP security authentication is configured for time synchronization in a cluster.

NOTE

This section applies only to physical machine clusters.

Prerequisites

- Manager has been installed and a third-party NTP server with the enabled authentication function has been appointed. Manager supports a maximum of two third-party NTP servers.
- You have obtained the NTP authentication key configured on the third-party NTP server from the server administrator.

Procedure

Step 1 Log in to the active management node as user **omm**.

This operation does not need to be performed on the standby management node.

Step 2 Run the following command to go to the related directory:

cd ${BIGDATA_HOME}/om-server/om/bin/tools/

Step 3 Run the following command to configure NTP security authentication for the cluster to access the third-party NTP server:

sh authenticateNtp.sh --ntp_server_ip *ip_address*

ip_address: indicates the IP address of the third-party NTP server. It does not support multiple IP addresses. That is, if there are multiple NTP servers, you need to run the command for each NTP server for authentication.

For example, run the following command to access the third-party NTP server with the IP address of **192.168.1.1**. The authentication key is **10 M b273290137CH**.

sh authenticateNtp.sh --ntp_server_ip 192.168.1.1

Enter the NTP authentication key as prompted, and press **Enter**. The authentication key index value cannot be **1**. You are advised to use other values.

```
Please input 192.168.1.1 authentication_code:
```


If the following information is displayed, the NTP security authentication is configured successfully:

```
Success to Authenticate ntp server to 192.168.1.1
```

----End

3.4.4 Enabling Two-Factor Authentication for a Cluster

Scenario

The RADIUS service is used to perform two-factor authentication for MRS Manager logins. By default, two-factor authentication is disabled. You can enable two-factor authentication to enhance cluster security.

Prerequisites

- Manager has been installed and is running properly.
- No critical alarms are generated for the MRS cluster.
- The RADIUS service has been configured.

Procedure

Step 1 Log in to the active management node as user **omm**.

Step 2 Run the following command to go to the related directory:

cd ${BIGDATA_HOME}/om-server/om/tools/

Step 3 Run the following command to generate the ciphertext of the shared key:

sh encrypt.sh --strSrc '*Shared key***' --encType 'AES256_CBC_IGNORE_KEY'**

Shared key is the one configured on the RADIUS server. For example:

sh encrypt.sh --strSrc '123456' --encType 'AES256_CBC_IGNORE_KEY'

```
[omm@kwephispra44947 tools]$ sh encrypt.sh --strSrc '123456' --encType 'AES256_CBC_IGNORE_KEY'
d2NjX2NyeXB0ATQxNDU1MzVGNDM0MjQzOzMyMzg0MjM0NDYzNjQzMzI0MjMzMzU0MzMwMzQzODQ1
NDE0MTQ0NDMzOTMxMzIzNTQ0MzIzMDQxMzgzMDM5MzE7OzMyMzUzMDMwO0MyMTY2RDU4NDk4N
UQ3RDMzMjlGMUI0OUQyRTcyOEU4O0Q4OTA4MkZGN0YyMzY1ODA7MzkzNDY1MzQzNTMxMzAzMjJEMzIz
MzY0NjMyRDM0NjUzNDM2MkQ2MTM0MzUzNzJENjIzNTY2MzkzNDYyMzg2MTM2NjU2MjY2Ow
```

Step 4 Run the following command to go to the related directory:

cd ${BIGDATA_OM_SERVER_HOME}/tomcat/webapps/web/WEB-INF/classes/config/

Step 5 Run the following command to open the **multifactor.properties** file:

vi multifactor.properties

Step 6 Change the value of **otp.switch** to **on**, set the **raduis.server.ip** parameter by referring to **Table 3-4**, and change the value of **radius.shared.secret** to the ciphertext of the shared key in **Step 3**. For example:

```
d2NjX2NyeXB0ATQxNDU1MzVGNDM0MjQzOzMyMzg0MjM0NDYzNjQzMzI0MjMzMzU0MzMwMzQzODQ1
NDE0MTQ0NDMzOTMxMzIzNTQ0MzIzMDQxMzgzMDM5MzE7OzMyMzUzMDMwO0MyMTY2RDU4NDk4N
UQ3RDMzMjlGMUI0OUQyRTcyOEU4O0Q4OTA4MkZGN0YyMzY1ODA7MzkzNDY1MzQzNTMxMzAzMjJEMzIz
MzY0NjMyRDM0NjUzNDM2MkQ2MTM0MzUzNzJENjIzNTY2MzkzNDYyMzg2MTM2NjU2MjY2Ow
```


Table 3-4 Parameters

Parameter	Description
otp.switch	Indicates whether to enable two-factor authentication. **on**: The authentication is enabled; **off**: The authentication is disabled. The default value is **off**.
auth.factor.type	Indicates the type of two-factor authentication. The default value is **radius**.
raduis.server.ip	Indicates the IP address of the RADIUS server used for two-factor authentication.
raduis.auth.port	Indicates the authentication port number of the RADIUS service. The default value is **1812**.
raduis.acct.port	Indicates the accounting port number of the RADIUS service. The default value is **1813**.
radius.shared.secret	Indicates the ciphertext of the two-factor authentication shared key.
radius.auth.protocol	Indicates the authentication protocol in use. The default value is **pap**.
ft.opt.acf.path	Other configuration, which is optional.

Step 7 (Skip this step if FusionInsight Manager is not installed on two management nodes.) Log in to the standby management node as user **omm** and perform **Step 2** to **Step 6**.

Step 8 Log in to the active management node as user **omm**.

Step 9 Run the following commands to restart the web service:

sh $OM_TOMCAT_HOME/bin/shutdown.sh

sh $OM_TOMCAT_HOME/bin/startup.sh

Step 10 Log in to MRS Manager. If OTP verification is displayed, two-factor authentication has been enabled.

----End

3.4.5 HFile and WAL Encryption

HFile and WAL Encryption

NOTICE

- Setting the HFile and WAL encryption mode to SMS4 or AES has a great impact on the system and will cause data loss in case of any misoperation. Therefore, this operation is not recommended.
- Batch data import using Bulkload does not support data encryption.

HFile and Write ahead log (WAL) in HBase are not encrypted by default. To encrypt them, perform the following operations.

Step 1 On any HBase node, run the following commands to create a key file as user **omm**:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.3.1/install/FusionInsight-HBase-*/hbase/bin/hbase-encrypt.sh *<path>/hbase.jks <type> <length> <alias>*

- */<path>/hbase.jks* indicates the path for storing the generated JKS file.
- *<type>* indicates the encryption type, which can be SMS4 or AES.
- *<length>* indicates the key length. SMS4 supports 16-bit and AES supports 128-bit.
- *<alias>* indicate the alias of the key file. When you create the key file for the first time, retain the default value **omm**.

For example, to generate an SMS4 encryption key, run the following command:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.3.1/install/FusionInsight-HBase-*/hbase/bin/hbase-encrypt.sh /home/hbase/conf/hbase.jks SMS4 16 omm

To generate an AES encryption key, run the following command:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.3.1/install/FusionInsight-HBase-*/hbase/bin/hbase-encrypt.sh /home/hbase/conf/hbase.jks AES 128 omm

NOTE

- To ensure operations can be successfully performed, the **<path>/hbase.jks** directory needs to be created in advance, and the cluster operation user must have the **rw** permission of this directory.
- After running the command, enter the same *<password>* four times. The password encrypted in **Step 3** is the same as the password in this step.

Step 2 Distribute the generated key files to the same directory on all nodes in the cluster and assign read and write permission to user **omm**.

NOTE

- Administrators need to select a safe procedure to distribute keys based on the enterprise security requirements.
- If the key files of some nodes are lost, repeat the step to copy the key files from other nodes.

Step 3 On FusionInsight Manager, set **hbase.crypto.keyprovider.parameters.encryptedtext** to the encrypted password. Set **hbase.crypto.keyprovider.parameters.uri** to the path and name of the key file.

- The format of **hbase.crypto.keyprovider.parameters.uri** is **jceks://** *<key_Path_Name>*.

  <key_Path_Name> indicates the path of the key file. For example, if the path of the key file is **/home/hbase/conf/hbase.jks**, set this parameter to **jceks:///home/hbase/conf/hbase.jks**.

- The format of **hbase.crypto.keyprovider.parameters.encryptedtext** is *<encrypted_password>*.

  <encrypted_password> indicates the encrypted password generated during the key file creation. The parameter value is displayed in ciphertext. Run the following command as user **omm** to obtain the related encrypted password on the nodes where HBase service is installed:

  sh ${BIGDATA_HOME}/FusionInsight_HD_8.3.1/install/FusionInsight-HBase-*/hbase/bin/hbase-encrypt.sh

  NOTE

  After running the command, you need to enter **<password>**. The password is the same as that entered in **Step 1**.

Step 4 On FusionInsight Manager, set **hbase.crypto.key.algorithm** to **SMS4** or **AES** to use SMS4 or AES for HFile encryption.

Step 5 On FusionInsight Manager, set **hbase.crypto.wal.algorithm** to **SMS4** or **AES** to use SMS4 or AES for WAL encryption.

Step 6 On FusionInsight Manager, set **hbase.regionserver.wal.encryption** to **true**.

Step 7 Save the settings and restart the HBase service for the settings to take effect.

Step 8 Create an HBase table through CLI or code and configure the encryption mode to enable encryption. **<type>** indicates the encryption type, and **d** indicates the column family.

- When you create an HBase table through CLI, set the encryption mode to SMS4 or AES for the column family.

  create '*<table name>*', {*NAME => 'd'*, ***ENCRYPTION => '****<type>*'}

- When you create an HBase table using code, set the encryption mode to SMS4 or AES by adding the following information to the code:

```
public void testCreateTable()
{
   String tableName = "user";
   Configuration conf = getConfiguration();
   HTableDescriptor htd = new HTableDescriptor(TableName.valueOf(tableName));

   HColumnDescriptor hcd = new HColumnDescriptor("d");
   //Set the encryption mode to SMS4 or AES.
   hcd.setEncryptionType("<type>");
   htd.addFamily(hcd);

   HBaseAdmin admin = null;
   try
   {
      admin = new HBaseAdmin(conf);
```



```
        if(!admin.tableExists(tableName))
        {
            admin.createTable(htd);
        }
    }
    catch (IOException e)
    {
        e.printStackTrace();
    }
    finally
    {
        if(admin != null)
        {
            try
            {
                admin.close();
            }
            catch (IOException e)
            {
                e.printStackTrace();
            }
        }
    }
}
```

Step 9 You can check whether the encryption configuration is successful by referring to **Verifying the Encryption Configuration**.

Step 10 If you have configured SMS4 or AES encryption by performing **Step 1** to **Step 7**, but do not set the related encryption parameter when creating the table in **Step 8**, the inserted data is not encrypted.

In this case, you can perform the following steps to encrypt the inserted data:

1. Run the **flush** command for the table to import the data in the memory to the HFile.

   flush *'<table_name>'*

2. Run the following commands to modify the table properties:

   disable *'<table_name>'*

   alter *'<table_name>'*,**NAME=>** *'<column_name>'*,**ENCRYPTION =>** *'<type>'*

   enable *'<table_name>'*

3. Insert a new data record and flush the table.

   NOTE

   A new data record must be inserted so that the HFile will generate a new HFile and the unencrypted data inserted previously will be rewritten and encrypted.

   put *'<table_name>'*, **'id2','f1:c1','value22222222222222222222222222222222'**

   flush *'<table_name>'*

4. Perform the following step to rewrite the HFile:

   major_compact'*<table_name>*'

NOTICE

During this step, the HBase table is disabled and cannot provide services. Exercise caution when you perform this step.

5. You can perform **Step 6** to check whether the encryption configuration is successful.

----End

Verifying the Encryption Configuration

NOTE

This operation can be performed only when test data can be written to an empty table.

Step 1 Log in to the node where the client is installed as the client installation user. Switch to the client installation directory, for example, **/opt/hadoopclient**.

cd /opt/hadoopclient

Step 2 Run the following command to set environment variables:

source bigdata_env

Step 3 Run the following command to authenticate the current user if the current cluster is a security cluster. The current user must have the permission to read and write HBase tables and the HDFS operation permission.

kinit *Component service user*

Run the following command to set the Hadoop username if the current cluster is in common mode:

export HADOOP_USER_NAME=hbase

Step 4 Run the following command to log in to the HBase client:

hbase shell

Run the following command to insert a new data record and flush the table to generate an HFile:

put *'<table_name>'*, **'id2','***d***:c1','value22222222222222222222222222222222'**

flush *'<table_name>'*

NOTE

- *<table_name>* indicates the table configured with SMS4 or AES encryption. For details about how to configure SMS4 or AES encryption, go to **Step 8**.
- *d* indicates the column family configured with SMS4 or AES encryption. For details about how to configure SMS4 or AES encryption, go to **Step 8**.

Step 5 Press **Ctrl**+**C** to exit the HBase client.

Step 6 Run the following command to view the directory where the HFile file generated in **Step 4** is stored:

hdfs dfs -ls

The file directory format is **/hbase/data/***<namespace_name>***/***<table_name>***/***<region_name>***/***<columnfamily_name>***/***<HFile_name>*.

NOTE

If *<namespace_name>* is not specified during HBase table creation, **default** is used by default.

Example:

```
/hbase/data/default/create_table/dd61b81b1ba1aad6513b9bdcfd8f871c/d/
aa6fe387b27443afaba40f5b584c1fa7
```

Step 7 Run the following command to view the HFile content:

hbase hfile -f *<HFile path>* **-p**

NOTE

<HFile path> indicates the directory where the HFile file is located in **Step 6**.

The error message "com. huawei.hadoop.hbase.io.crypto.CryptoRuntimeException" will be displayed in the command output. However, the **HBase shell** can still read the table data, indicating that the encryption configuration is successful.

----End

Modifying a Key File

NOTICE

Modifying a key file has a great impact on the system and will cause data loss in case of any misoperation. Therefore, this operation is not recommended.

During the **HFile and WAL Encryption** operation, the related key file must be generated and its password must be set to ensure system security. After a period of running, you can replace the key file with a new one to encrypt HFile and WAL.

Step 1 Run the following command to generate a new key file as user **omm**:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.3.1/install/FusionInsight-HBase-*/hbase/bin/hbase-encrypt.sh *<path>/hbase.jks <type> <length> <alias-new>*

- *<path>/hbase.jks*: indicates the path for storing the generated **hbase.jks** file. The path and file name must be consistent with those of the key file generated in **HFile and WAL Encryption**.
- *<alias-new>*: indicates the alias of the key file. The alias must be different with that of the old key file.
- *<type>*: indicates the encryption type, which can be SMS4 or AES.
- *<length>* indicates the key length. SMS4 supports 16-bit and AES supports 128-bit.

For example, to generate an SMS4 encryption key, run the following command:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.3.1/install/FusionInsight-HBase-*/hbase/bin/hbase-encrypt.sh /home/hbase/conf/hbase.jks SMS4 16 omm_new

To generate an AES encryption key, run the following command:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.3.1/install/FusionInsight-HBase-*/hbase/bin/hbase-encrypt.sh /home/hbase/conf/hbase.jks AES 128 omm_new

NOTE

- To ensure operations can be successfully performed, the **<path>/hbase.jks** directory needs to be created in advance, and the cluster operation user must have the **rw** permission of this directory.
- After running the command, you need to enter the same *<password>* for three times. This password is the password of the key file. You can use the password of the old file without any security risk.

Step 2 Distribute the generated key files to the same directory on all nodes in the cluster and assign read and write permission to user **omm**.

NOTE

Administrators need to select a safe procedure to distribute keys based on the enterprise security requirements.

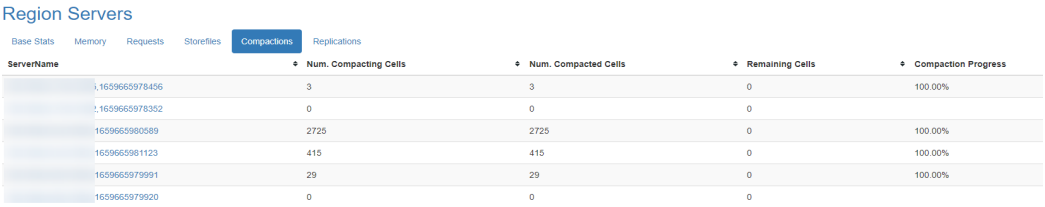
Step 3 On the HBase service configuration page of FusionInsight Manager, add custom configuration items, set **hbase.crypto.master.key.name** to **omm_new**, set **hbase.crypto.master.alternate.key.name** to **omm**, and save the settings.

Step 4 Restart the HBase service for the configuration to take effect.

Step 5 In HBase shell, run the **major compact** command to generate the HFile file based on the new encryption algorithm.

major_compact *'<table_name>'*

Step 6 You can view the major compact progress from the HMaster web page.



Step 7 When all items in **Compaction Progress** reach **100%** and those in **Remaining KVs** are **0**, run the following command as user **omm** to destroy the old key file:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.3.1/install/FusionInsight-HBase-*/hbase/bin/hbase-encrypt.sh *<path>/hbase.jks <alias-old>*

- *<path>/hbase.jks*: indicates the path for storing the generated **hbase.jks** file. The path and file name must be consistent with those of the key file generated in **HFile and WAL Encryption**.
- *<alias-old>*: indicates the alias of the old key file to be deleted.

For example:

sh ${BIGDATA_HOME}/FusionInsight_HD_8.3.1/install/FusionInsight-HBase-*/hbase/bin/hbase-encrypt.sh /home/hbase/conf/hbase.jks omm

NOTE

To ensure operations can be successfully performed, the **<path>/hbase.jks** directory needs to be created in advance, and the cluster operation user must have the **rw** permission of this directory.

Step 8 Repeat **Step 2** and distribute the updated key files again.

Step 9 Delete the HBase self-defined configuration item **hbase.crypto.master.alternate.key.name** added in **Step 3** from FusionInsight Manager.

Step 10 Repeat **Step 4** for the configuration take effect.

----End

3.4.6 Configuring Hadoop Security Parameters

Configuring Security Channel Encryption

The channels between components are not encrypted by default. You can set the following parameters to configure security channel encryption.

To modify parameters, log in to FusionInsight Manager, choose **Cluster** > **Services** > *Service name*, click **Configurations** then **All Configurations**, and enter a parameter name in the search box.

NOTE

Restart corresponding services for the modification to take effect after you modify configuration parameters.

Table 3-5 Parameter description

Service	Parameter	Description	Default Value
HBase	hbase.rpc.protection	Indicates whether the HBase channels, including the remote procedure call (RPC) channels for HBase clients to access the HBase server and the RPC channels between the HMaster and RegionServer, are encrypted. If this parameter is set to **privacy**, the channels are encrypted and the authentication, integrity, and privacy functions are enabled. If this parameter is set to **integrity**, the channels are not encrypted and only the authentication and integrity functions are enabled. If this parameter is set to **authentication**, the channels are not encrypted, only packets are authenticated, and integrity and privacy are not required. **NOTE** The privacy mode encrypts transmitted content, including sensitive information such as user tokens, to ensure the security of the transmitted content. However, this mode has great impact on performance. Compared with the other two modes, this mode reduces read/write performance by about 60%. Modify the configuration based on the enterprise security requirements. The configuration items on the client and server must be the same.	• Security mode: privacy • Normal mode: authentication
HDFS	dfs.encrypt.data.transfer	Indicates whether the HDFS data transfer channels and the channels for clients to access HDFS are encrypted. The HDFS data transfer channels include the data transfer channels between DataNodes and the Data Transfer (DT) channels for clients to access DataNodes. The value **true** indicates that the channels are encrypted. The channels are not encrypted by default.	false

Service	Parameter	Description	Default Value
HDFS	dfs.encrypt.data.transfer.algorithm	Indicates whether the HDFS data transfer channels and the channels for clients to access HDFS are encrypted. This parameter is valid only when **dfs.encrypt.data.transfer** is set to **true**. The default value is **3des**, indicating that 3DES algorithm is used to encrypt data. The value can also be set to **rc4**. However, to avoid security risks, you are not advised to set the parameter to this value.	3des
HDFS	hadoop.rpc.protection	Indicates whether the RPC channels of each module in Hadoop are encrypted. The channels include: • RPC channels for clients to access HDFS • RPC channels between modules in HDFS, for example, between DataNode and NameNode • RPC channels for clients to access Yarn • RPC channels between NodeManager and ResourceManager • RPC channels for Spark to access Yarn and HDFS • RPC channels for MapReduce to access Yarn and HDFS • RPC channels for HBase to access HDFS The default value is **privacy**, indicating encrypted transmission. The value **authentication** indicates that transmission is not encrypted. **NOTE** You can set this parameter on the HDFS component configuration page. The parameter setting is valid globally, that is, the setting of whether the RPC channel is encrypted takes effect on all modules in Hadoop.	• Security mode: **privacy** • Normal mode: **authentication**

Setting the Maximum Number of Concurrent Web Connections

To ensure web server reliability, new connections are rejected when the number of user connections reaches a specific threshold. This prevents DDOS attacks and

service unavailability caused by too many users accessing the web server at the same time.

To modify parameters, log in to FusionInsight Manager, choose **Cluster** > **Services** > *Service name*, click **Configurations** then **All Configurations**, and enter a parameter name in the search box.

Table 3-6 Parameter description

Service	Parameter	Description	Default Value
HDFS/Yarn	hadoop.http.server.MaxRequests	Specifies the maximum number of concurrent web connections of each component.	2000
Spark	spark.connection.maxRequest	Specifies the maximum number of request connections of JobHistory.	5000

3.4.7 Configuring an IP Address Whitelist for Modification Allowed by HBase

If the Replication function is enabled for HBase clusters, a protection mechanism for data modification is added on the standby HBase cluster to ensure data consistency between the active and standby clusters. Upon receiving an RPC request for data modification, the standby HBase cluster checks the permission of the user who sends the request (only HBase manage users have the modification permission). Then it checks the validity of the source IP address of the request. Only modification requests from IP addresses in the white list are accepted. The IP address white list is configured by the **hbase.replication.allowedIPs** item.

Log in to FusionInsight Manager and choose **Cluster** > **Services** > **HBase**. Click **Configurations** and enter the parameter name in the search box.

Table 3-7 Parameter description

Parameter	Description	Default Value
hbase.replication.allowedIPs	Allows replication request processing from configured IP addresses only. It supports comma separated regex patterns. Each pattern can be any of the following: • Regex pattern Example: 10.18.40.*, 10.18.*, 10.18.40.11 • Range pattern (Range can be specified only in the last octet) Example: 10.18.40.[10-20] If this item is empty (default value), the white list contains only the IP address of the RegionServer of the cluster, indicating that only modification requests from the RegionServer of the standby HBase cluster are accepted.	N/A

3.4.8 Updating a Key for a Cluster

Scenario

When a cluster is installed, an encryption key is generated automatically by the system so that the security information in the cluster (such as all database user passwords and key file access passwords) can be stored in encryption mode. After the cluster is installed, if the original key is accidentally disclosed or a new key is required, you can manually update the key.

Impact on the System

- After a cluster key is updated, a new key is generated randomly in the cluster. This key is used to encrypt and decrypt the newly stored data. The old key is not deleted, and it is used to decrypt data encrypted using the old key. After security information is modified, for example, a database user password is changed, the new password is encrypted using the new key.
- When a key is updated for a cluster, the cluster must be stopped and cannot be accessed.

Prerequisites

- You have obtained the IP addresses of the active and standby management nodes. For details, see **Logging In to the Management Node**.
- You have stopped the upper-layer service applications that depend on the cluster.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 In the upper right corner of **Homepage**, click **Stop**. In the dialog box displayed, enter the password of the current user for identity confirmation.

and click **OK**. Wait for a while until a message indicating that the operation is successful is displayed.

Step 3 Log in to the active management node as user **omm**.

Step 4 Run the following command to disable logout upon timeout:

TMOUT=0

NOTE

After the operations in this section are complete, run the **TMOUT=***Timeout interval* command to restore the timeout interval in a timely manner. For example, **TMOUT=600** indicates that a user is logged out if the user does not perform any operation within 600 seconds.

Step 5 Run the following command to go to the related directory:

cd ${BIGDATA_HOME}/om-server/om/tools

Step 6 Run the following command to update the cluster key:

sh updateRootKey.sh

Enter **y** as prompted.

```
The root key update is a critical operation.
Do you want to continue?(y/n):
```

If the following information is displayed, the key is updated successfully.

```
Step 4-1: The key save path is obtained successfully.
...
Step 4-4: The root key is sent successfully.
```

Step 7 In the upper right corner of **Homepage**, click **Start**.

In the displayed dialog box, click **OK**. Wait until a message is displayed, indicating that the startup is successful.

----End

3.4.9 Changing the Cluster Encryption Mode

Scenario

This section describes how to change the encryption mode of a cluster.

Impact on the System

When changing the encryption mode of a cluster, the cluster and OMS node are stopped and cannot be accessed.

Prerequisites

The upper-layer applications depending on the cluster are stopped.

Procedure

Step 1 Log in to FusionInsight Manager as user **admin**.

Step 2 In the upper right corner of **Homepage**, click **Stop**. In the dialog box displayed, enter the password of the current user for identity confirmation.

and click **OK**. Wait for a while until a message indicating that the operation is successful is displayed.

Step 3 Log in to the active management node as user **root** and run the following command to switch to user **omm**:

su - omm

Step 4 Run the following command to check the current encryption mode of the cluster (that is, the value of the **defaultAlgorithm** parameter in the **scc.conf** file):

cat $BIGDATA_COMMON/securityforscc/config/scc.conf

For example, the following information indicates that the current cluster is encrypted using the general encryption algorithm.

```
.....
defaultAlgorithm=AES256_GCM
.....
```

Step 5 Run the following commands to change the cluster encryption mode, for example, to SMCompatible:

cd $CONTROLLER_HOME/tools

bash updateSysSecretMain.sh -o update -a *SMCompatible*

For details about the parameters of the script for changing the encryption mode, see **Reference Information**.

The cryptographic algorithm is successfully changed if the following information is displayed:

```
start to pre-action(update)
end to pre-action(update)
Operations(update) need to be performed on 3 nodes in the cluster.
start to execute action(update) on node[No:1, ip:192.168.43.43, nodeType:oms-node-active]
end to execute action(update) on node[No:1, ip:192.168.43.43, nodeType:oms-node-active]
.....
start to post-action(update)
end to post-action(update)
execute action(update) success.
```

Step 6 Run the following command to check view cluster encryption mode:

cat $BIGDATA_COMMON/securityforscc/config/scc.conf

```
.....
defaultAlgorithm=SM4_CTR
.....
```


Step 7 In the upper right corner of **Homepage**, click **More** and select **Synchronize Configurations**. In the dialog box displayed, click **OK** to synchronize configurations for the current cluster. Wait until the synchronization is complete.

Step 8 Click **Start**. In the displayed dialog box, click **OK**. Wait until a message is displayed indicating that the startup is successful.

Step 9 Check whether the cluster is successfully started and all services are running properly.

- If yes, go to **Step 10**.
- If no, go to **Step 11**.

Step 10 After the cluster is started and services are running properly, run the following commands on the active management node of the cluster to delete the files related to the old key:

cd $CONTROLLER_HOME/tools

bash updateSysSecretMain.sh -o commit

The operation is successful if the following information is displayed:

```
Operations(commit) need to be performed on 3 nodes in the cluster.
start to execute action(commit) on node[No:1, ip:192.168.43.43, nodeType:oms-node-active]
end   to execute action(commit) on node[No:1, ip:192.168.43.43, nodeType:oms-node-active]
......
execute action(commit) success.
```

Step 11 If the cluster fails to be started or the service running status is abnormal, run the following commands on the active management node of the cluster to roll back to the state before the encryption mode of the cluster is changed. If the rollback fails, contact technical support.

cd $CONTROLLER_HOME/tools

bash updateSysSecretMain.sh -o rollback

The operation is successful if the following information is displayed:

```
start to pre-action(rollback)
end to pre-action(rollback)
Operations(rollback) need to be performed on 3 nodes in the cluster.
start to execute action(rollback) on node[No:1, ip:192.168.43.43, nodeType:oms-node-active]
end   to execute action(rollback) on node[No:1, ip:192.168.43.43, nodeType:oms-node-active]
......
start to post-action(rollback)
end to post-action(rollback)
execute action(rollback) success.
```

Run the following command to submit the rollback operation:

bash updateSysSecretMain.sh -o commit

The operation is successful if the following information is displayed:

```
Operations(commit) need to be performed on 3 nodes in the cluster.
start to execute action(commit) on node[No:1, ip:192.168.43.43, nodeType:oms-node-active]
end   to execute action(commit) on node[No:1, ip:192.168.43.43, nodeType:oms-node-active]
......
execute action(commit) success.
```

----End

Reference Information

The following describes the parameters of the script for changing the encryption mode.

```
help:
  parameters:
     -o: Operation Type, Mandatory parameters, Enumerated Value: update | commit | rollback
     -a: Algorithm Type, Optional parameters(Required only for update operation), Enumerated Value:
generalCipher | SMCompatible | SMOnly
  usage:
     updateSysSecretMain.sh -o [ update | commit | rollback ] | [ -a [ generalCipher | SMCompatible |
SMOnly ] ]
```

- **-o**: indicates the supported operations for changing the encryption mode of a cluster key, including the update, rollback, and commit operations. The update or rollback operation is followed by a commit operation, which is used to submit the current operation result.
- **-a**: indicates the type of an encryption mode. The update operation supports the following key modes:
  - **generalCipher**: indicates that the general encryption mode is used.
  - **SMCompatible/SMOnly**: indicates that the national encryption mode is used.

3.4.10 Hardening the LDAP

Configuring the LDAP Firewall Policy

In the cluster adopting the dual-plane networking, the LDAP is deployed on the service plane. To ensure the LDAP data security, you are advised to configure the firewall policy in the cluster to disable relevant LDAP ports.

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **Cluster** > **Services** > **LdapServer** and click **Configurations**.

Step 3 Check the value of **LDAP_SERVER_PORT**, which is the service port of LdapServer.

Step 4 To ensure data security, configure the firewall policy for the whole cluster to disable the LdapServer port based on the customer's firewall environment.

----End

Enabling the LDAP Audit Log Output

Users can set the audit log output level of the LDAP service and output audit logs in a specified directory, for example, **/var/log/messages**. The logs output can be used to check user activities and operation commands.

NOTE

If the function of LDAP audit log output is enabled, massive logs are generated, affecting the cluster performance. Exercise caution when enabling this function.

Step 1 Log in to any LdapServer node.

Step 2 Run the following command to edit the **slapd.conf.consumer** file, and set the value of **loglevel** to **256** (you can run the **man slapd.conf** command on the OS to view the log level definition).

cd ${BIGDATA_HOME}/FusionInsight_BASE_8.3.1/install/FusionInsight-ldapserver-2.7.0/ldapserver/local/template

vi slapd.conf.consumer

```
...
pidfile         [PID_FILE_SLAPD_PID]
argsfile        [PID_FILE_SLAPD_ARGS]
loglevel   256
...
```

Step 3 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **LdapServer**. Click **More** and select **Restart Service**. In the dialog box displayed, verify the current user identity, and restart the service.

----End

3.4.11 Configuring Kafka Data Encryption During Transmission

Scenario

Data between the Kafka client and the broker is transmitted in plain text. The Kafka client may be deployed in an untrusted network, exposing the transmitting data to leakage and tampering risks.

Procedure

The channel between components is not encrypted by default. You can set the following parameters to enable security channel encryption.

To modify parameters, log in to FusionInsight Manager, choose **Cluster** > **Services** > **Kafka**, and click **Configurations** then **All Configurations**. Enter a parameter name in the search box.

NOTE

After the configuration, restart the corresponding service for the settings to take effect.

Table 3-8 describes the parameters related to transmission encryption on the Kafka server.

Table 3-8 Parameters relevant to Kafka data encryption during transmission

Parameter	Description	Default Value
ssl.mode.enable	Indicates whether to enable the Secure Sockets Layer (SSL) protocol. If this parameter is set to **true**, services relevant to the SSL protocol are started during the broker startup.	false

Parameter	Description	Default Value
security.inter.broker.protocol	Indicates communication protocol between brokers. The communication protocol can be PLAINTEXT, SSL, SASL_PLAINTEXT, or SASL_SSL.	SASL_PLAINTEXT

The SSL protocol can be configured for the server or client to encrypt transmission and communication only after **ssl.mode.enable** is set to **true** and broker enables the **SSL** and **SASL_SSL** protocols.

For details about the access protocol, see " Protocol Description for Accessing Kafka" in **Component Operation Guide > Using Kafka > Safety Instructions on Using Kafka**.

For details about how to configure SSL encrypted transmission on the client, see "Security Mode" > "Kafka Development Guide" > "More Information" > "External APIs" > "SSL Encryption Function Used by a Client" in *MapReduce Service (MRS) 3.3.1-LTS Developer Guide (for Huawei Cloud Stack 8.3.1) in the MapReduce Service (MRS) 3.3.1-LTS Usage Guide (for Huawei Cloud Stack 8.3.1)*.

3.4.12 Configuring HDFS Data Encryption During Transmission

Configuring HDFS Security Channel Encryption

The channel between components is not encrypted by default. You can set parameters to enable security channel encryption.

To modify parameters, log in to FusionInsight Manager, choose **Cluster** > **Services** > **HDFS**, and click **Configurations** then **All Configurations**. Enter a parameter name in the search box.

NOTE

After the configuration, restart the corresponding service for the settings to take effect.

Table 3-9 Parameters

Configuration Item	Description	Default Value
hadoop.rpc.protection	**NOTICE** • The setting takes effect only after the service is restarted. Rolling restart is not supported. • After the setting, you need to download the client configuration file again. Otherwise, HDFS cannot provide the read and write services. • After the setting, you need to restart the executor. Otherwise, the job management and file management functions on the console become unavailable. Indicates whether the RPC channels of each module in Hadoop are encrypted. The channels include: • RPC channels for clients to access HDFS • RPC channels between modules in HDFS, for example, between DataNode and NameNode • RPC channels for clients to access Yarn • RPC channels between NodeManager and ResourceManager • RPC channels for Spark to access Yarn and HDFS • RPC channels for MapReduce to access Yarn and HDFS • RPC channels for HBase to access HDFS • RPC channels for CDL to submit tasks **NOTE** The setting takes effect globally, that is, the encryption attribute of the RPC channel of each module in the Hadoop takes effect.	• Security mode: privacy • Normal mode: authentication **NOTE** • **authentication**: indicates that only authentication is required. • **integrity**: indicates that authentication and consistency check need to be performed. • **privacy**: indicates that authentication, consistency check, and encryption need to be performed.

Configuration Item	Description	Default Value
dfs.encrypt.data.transfer	Indicates whether the HDFS data transfer channels and the channels for clients to access HDFS are encrypted. The HDFS data transfer channels include the data transfer channels between DataNodes and the Data Transfer (DT) channels for clients to access DataNodes. The value **true** indicates that the channels are encrypted. The channels are not encrypted by default. **NOTE** • This parameter is valid only when **hadoop.rpc.protection** is set to **privacy**. • If a large amount of service data is transmitted, enabling encryption by default severely affects system performance. • When Router-based Federation is used, this encryption feature cannot be enabled. ECS/BMS clusters do not support Federation. • If data transmission encryption is configured for one cluster in the trusted cluster, the same data transmission encryption must be configured for the peer cluster.	false
dfs.encrypt.data.transfer.algorithm	Indicates the algorithm to encrypt the HDFS data transfer channels and the channels for clients to access HDFS. This parameter is valid only when **dfs.encrypt.data.transfer** is set to **true**. **NOTE** The default value is **3des**, indicating that 3DES algorithm is used to encrypt data. The value can also be set to **rc4**. However, to avoid security risks, you are not advised to set the parameter to this value.	3des

Configuration Item	Description	Default Value
dfs.encrypt.data.transfer.cipher.suites	This parameter can be left empty or set to **AES/CTR/NoPadding** to specify the cipher suite for data encryption. If this parameter is not specified, the encryption algorithm specified by **dfs.encrypt.data.transfer.algorithm** is used for data encryption. The default value is **AES/CTR/NoPadding**.	AES/CTR/NoPadding
dfs.data.transfer.protection	Whether to encrypt the RPC channel used by the HDFS client to read and write data. There are three encryption methods: • **authentication**: only authentication is required. • **integrity**: authentication and consistency check are required. • **privacy**: authentication, consistency check, and encryption are required. **NOTICE** After the configuration is modified, you need to restart the HDFS service and its upper-layer services. Rolling restart is not supported. Services will be interrupted during the restart. Exercise caution when performing this operation.	-

3.4.13 Configuring HetuEngine Data Encryption During Transmission

Scenario

This section describes how to configure HTTPS encryption for communication between nodes in a cluster and configure a whitelist for accessing HSConsole to enhance security.

NOTE

You are advised to use the secure HTTPS protocol. Risks exist if you use an insecure protocol.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **HetuEngine**. Click **Configurations** then **All Configurations**. Enter the parameter name in the search box.

NOTE

After the configuration, restart the corresponding service for the settings to take effect.

Table 3-10 Security configuration

Parameter	Description	Default Value
internal-communication.https.required	Whether communication between nodes in a cluster requires HTTPS encryption. If this option is enabled, the query performance may deteriorate.	true **NOTE** If this parameter is set to **false**, **http-server.http.enabled** must be enabled.
referer.whitelist	Whitelist of web request headers that are allowed to access the HSconsole. Use semicolons (;) to separate multiple whitelists, for example, "https://192.168.1.2:25000:*;https://192.168.1.3:25001:*".	N/A
http-server.https.enabled	Whether to enable HTTPS access for HetuEngine Computer Cluster.	true **NOTICE** If it is set to false, the HTTP protocol is used, please ensure that HetuEngine Compute Instance works in secure context.

----End

3.4.14 Configuring RTD Data Encryption During Transmission

Scenario

Configure RTD security channel encryption to enhance security.

Procedure

Step 1 Log in to FusionInsight Manager as a service user and choose **Cluster** > **Services** > **MOTService**. Click **Configurations** then **All Configurations**.

Step 2 Click **MOTServer(Role)**, select **Security**, and check whether the value of the following parameter is **true**. If not, change the value to **true** and save the settings.

Table 3-11 Parameters

Parameter	Description	Example Value
REQUIRE_SSL	The server forcibly requires the SSL connection.	true

Step 3 Click **Dashboard**, click **More**, and select **Restart Service** to restart MOTService as prompted.

Step 4 When creating a MOT cluster tenant, select **Enable SSL**.

1. Log in to FusionInsight Manager as a service user and choose **Cluster** > **Services** > **RTDService**.
2. Click the link next to **RTD WebUI** to access the RTD web UI.
3. Choose **System** > **Tenant Management**. Click **Add**, enter a tenant name, and set **DB Type** to **MOT**.
4. Click **MOT Cluster**, select **Enable SSL** (set other parameters based on site requirements), and click **OK**.

----End

3.4.15 Configuring Spark Data Encryption During Transmission

Scenario

Configure encryption for Spark security channels to enhance security.

Procedure

To modify parameters, log in to FusionInsight Manager, click **Cluster** and choose **Services** > **Spark**. On the displayed page, click **Configurations** and click **All Configurations**. Enter a parameter name in the search box.

NOTE

After the configuration, restart the corresponding service for the settings to take effect.

Table 3-12 Parameters

Parameter	Description	Default Value
spark.authenticate	Whether to enable Spark internal security authentication	Security mode: **true** Normal mode: **false**
spark.authenticate.enableSaslEncryption	Whether to enable encrypted communication based on Simple Authentication and Security Layer (SASL)	Security mode: **true** Normal mode: **false**
spark.network.crypto.enabled	Whether to enable RPC encryption based on Advanced Encryption Standard (AES)	false
spark.network.sasl.serverAlwaysEncrypt	Whether to disable unencrypted connections for ports with SASL authentication enabled	false
spark.network.crypto.keyLength	Length of the encryption key to be generated	256
spark.network.crypto.keyFactoryAlgorithm	Algorithm used to generate the encryption key	PBKDF2WithHmacSHA256
spark.io.encryption.enabled	Whether to enable local disk I/O encryption	Security mode: **true** Normal mode: **false**
spark.io.encryption.keygen.algorithm	Algorithm used to generate the I/O encryption key	HmacSHA256
spark.io.encryption.keySizeBits	Size of an I/O encryption key, in bits	256
spark.ssl.ui.enabled	Whether to enable Secure Sockets Layer (SSL) authentication for the web UI connection	Security mode: **true** Normal mode: **false**

3.4.16 Configuring IoTDB Data Encryption During Transmission

Scenario

In security scenarios, IoTDB data transmission encryption is required to ensure data security.

NOTE

Data transmission encryption affects performance. Therefore, you are not advised to enable this function in scenarios that require high performance.

Prerequisites

- Users of service components have been created by the system administrator as required. For details, see **Creating a User**. In security mode, machine-to-machine users have downloaded the keytab file. For details about how to download the keytab files, see **Exporting an Authentication Credential File**. A human-machine user must change the password upon the first login.
- The IoTDB client has been installed in a directory, for example, **/opt/hadoopclient**. Choose **Cluster** > **Services** > **IoTDB** then choose **More** > **Download Client**. In the dialog box that is displayed, select **Save to Path**. The generated file is saved to the **/tmp/FusionInsight-Client** directory on the active management node by default.

Procedure

Step 1 Perform operations on the server.

1. Log in to FusionInsight Manager and choose **Cluster** > **Services** > **IoTDB**. Click the **Configurations** tab.
2. Modify the following parameters:
   - Search for **SSL_ENABLE** in the upper right corner of the page and change its value to **true**.
   - Search for **iotdb_server_kerberos_qop** in the upper right corner of the page and change its value to **auth-conf**.
3. Click **Save**. In the **Save Configuration** dialog box, click **OK**.

   When **Operation succeeded** is displayed, click **Finish**.
4. Click the **Dashboard** tab. Choose **More** > **Restart Service**. Wait until the service is restarted.

Step 2 Perform operations on the client.

1. Log in to the active management node as user **root** and run the following command to switch to the client installation directory, for example, **/opt/hadoopclient**:

   cd /opt/hadoopclient
2. Run the following command to configure environment variables:

   source bigdata_env

3. (Optional) Run the following command to authenticate the current user if Kerberos authentication is enabled for the cluster. If Kerberos authentication is not enabled, skip this step.

   kinit *Component service user*

4. Run the following commands to generate the **truststore.jks** file using the **ca.crt** certificate file in the root directory of the client:

   cd /tmp/FusionInsight-Client/FusionInsight_Cluster_*_IoTDB_ClientConfig

   keytool -noprompt -import -alias myservercert -file ca.crt -keystore truststore.jks

5. Run the following command to copy the generated **truststore.jks** file to the client installation directory, for example, **/opt/hadoopclient/IoTDB/iotdb/conf**.

   cp truststore.jks /opt/hadoopclient/IoTDB/iotdb/conf

6. Run the following command to switch to the directory where the IoTDB client running script is stored:

   cd /opt/hadoopclient/IoTDB/iotdb/sbin

7. Run the **vim start-cli.sh** command to compile the **start-cli.sh** script and add **iotdb_ssl_truststore=/opt/hadoopclient/IoTDB/iotdb/conf/truststore.jks** and **iotdb_ssl_enable=true** to the startup parameter **exec "$JAVA" -cp "$JAVA_CLASSPATH" "$JAVA_MAIN_CLASS" $PARAMETERS** in the script.

   ```
   exec "$JAVA" -Diotdb_ssl_truststore=/opt/hadoopclient/IoTDB/iotdb/conf/truststore.jks -Diotdb_ssl_enable=true -cp "$JAVA_CLASSPATH" "$JAVA_MAIN_CLASS" $PARAMETERS
   ```

8. Run the following command to log in to the client. If the login is successful, data transmission encryption is enabled for IoTDB.

   ./start-cli.sh -h *Service IP address of node where the IoTDBServer instance is located* **-p** *IoTDBServer RPC port*

   NOTE

   - You can also log in to the client by running the **./start-cli.sh -h** *Service IP address of the node where the IoTDBServer instance is located* **-p** *IoTDBServer RPC port* **-u** *Service username* **-pw** *Service user password* command.
   - To view the service IP address of the node where the IoTDBServer instance is located, log in to FusionInsight Manager, choose **Cluster** > **Services** > **IoTDB**, and click the **Instance** tab.
   - The default RPC port is **22260**. To obtain the port number, choose **Cluster** > **Services** > **IoTDB**, click **Configurations** then **All Configurations**, and search for **IOTDB_SERVER_RPC_PORT**.

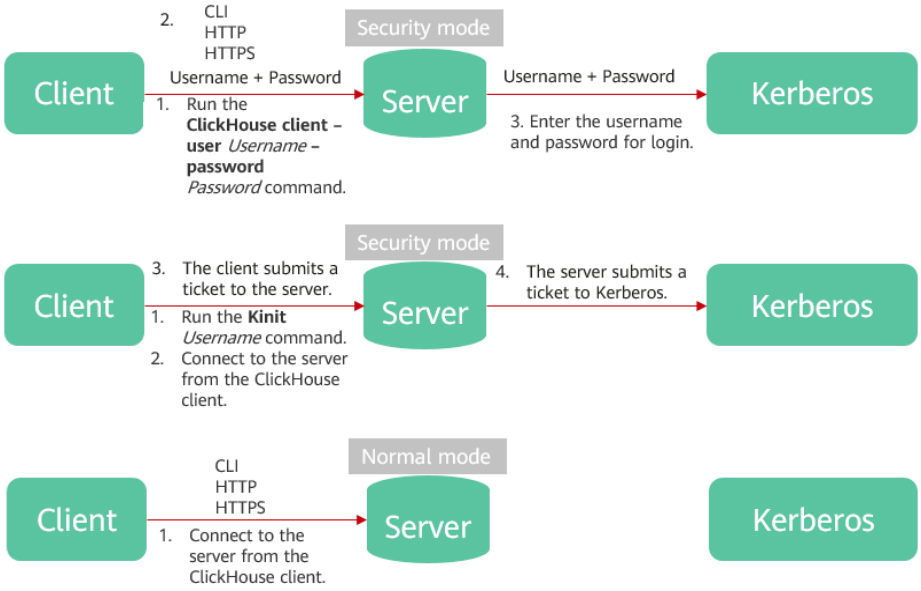
----End

3.4.17 ClickHouse Security Hardening

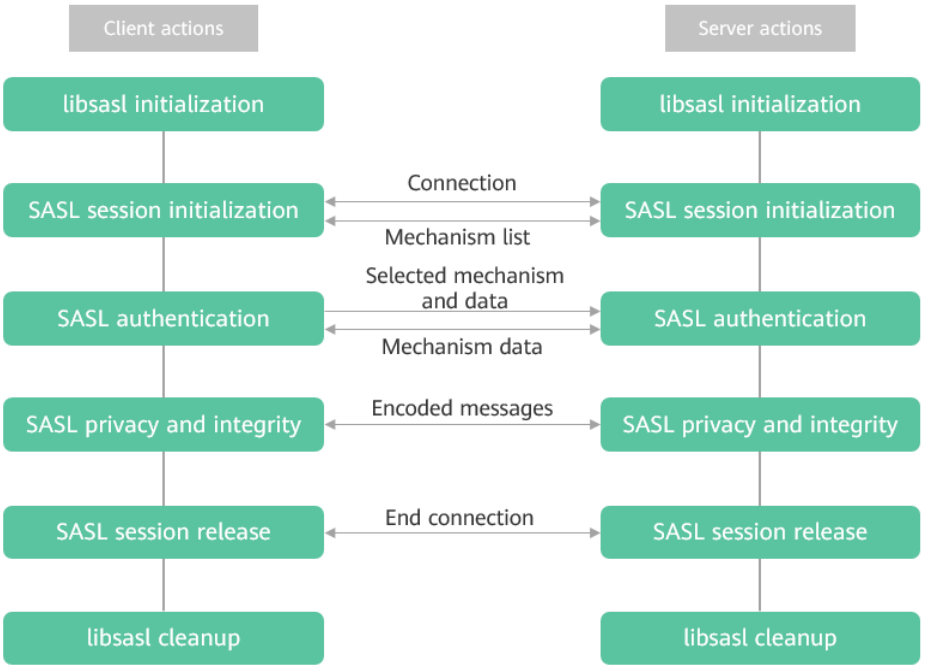
Authentication and Encryption

The authentication system of ClickHouse is as follows:

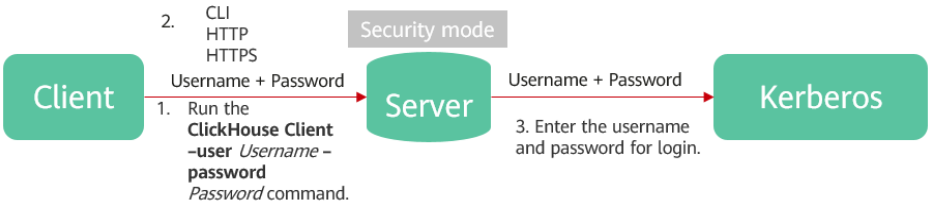
Figure 3-2 ClickHouse system authentication mode



- The normal mode does not require authentication. You can use the built-in default user to log in to the system without authentication.
- The kinit authentication mode of the client in security mode uses the sasl authentication mode. The implementation principle is as follows:



- The client in security mode is compatible with the community version. Kerberos authentication is performed only on the server.



- ClickHouse supports permission management for the following objects:

Resource	Permission
Database	CREATE
Table/View	SELECT/INSERT
Admin	ALL

NOTE

ClickHouse does not support disabling security authentication on a cluster in security mode.

Encrypted Channel

ClickHouse enhances usability based on the open-source community version. By default, clusters in security mode use TCP and HTTP channels encrypted by OpenSSL.

Security Hardening

- Encoding rules

  Description: The same encoding mode is used on the web service client and server to prevent garbled characters and to implement input verification.

  Security hardening: Response messages of web servers are encoded using UTF-8.

- IP address whitelist filtering supported for management users

  Description: IP address whitelist filtering is used to prevent unauthorized clients from logging in to the system.

  Security hardening: External nodes are not allowed to access the ClickHouse client as a management user.

- URL injection attack

  Description: A customized UI of data migration is used to prevent URL injection attacks.

  Security hardening: URL and path validity checks are implemented.

- SQL injection attack

  Description: ClickHouse prevents SQL injection attacks.

  Security hardening: SQL statements are precompiled.

- Log injection attack

  Description: Log injection must be prevented to avoid security information leakage.

Security hardening: Privacy information is encrypted to prevent sensitive stack information from being recorded in logs.

- DDoS attack

  Description: ClickHouse prevents service interruption or exceptions caused by DDoS attacks.

  Security hardening: The number of connections is configurable. The default value is 4096.

- Anti-repudiation

  Description: Audit logs are recorded.

  Security hardening: DDL operations such as write, permission granting and revoking, and data migration are audited.

3.4.18 Configuring ZooKeeper SSL

Scenario

By default, SSL channel encryption transmission is disabled between the ZooKeeper client and server and between instances on the server. This section describes how to enable the ZooKeeper channel encryption transmission.

Impact on the System

- When SSL channel encryption transmission is enabled on the ZooKeeper server, the performance deteriorates.
- When SSL channel encryption transmission is enabled on the ZooKeeper server, ZooKeeper and dependent upper-layer components need to be restarted. During the restart, services are unavailable.
- To enable SSL channel encryption transmission on the ZooKeeper server, you need to download the client again.
- If SSL channel encryption transmission is enabled for ZooKeeper, rolling restart is not supported.
- Guardian does not support SSL for ZooKeeper.

Procedure

Step 1 Log in to FusionInsight Manager, click **Cluster** and choose **Services** > **ZooKeeper**. On the displayed page, click **Configurations** and click **All Configurations**.

Step 2 Enter the parameter name in the search box, and change the value as follows:

Table 3-13 Security configuration item

Parameter	Description	Default Value	New Value
ssl.enabled	Whether to enable SSL communication encryption.	false	true

Step 3 After the modification is complete, click **Save** and then click **OK**.

Step 4 Click **Cluster** and choose **Services** > **ZooKeeper**. On the ZooKeeper service page, choose **More** > **Restart Service**, enter the password for authentication, and confirm the operation impact on the **Restart Service** page.

You can select **Restart upper-layer services**. During the restart of all affected components, services will be unavailable. Exercise caution when performing this operation.

Step 5 Click **OK** and wait until the services are restarted successfully.

Step 6 Choose **Cluster** > **Active/Standby Cluster DR** to check whether active/standby DR is configured for the current cluster.

- If yes, go to **Step 7**.
- If no, no further action is required.

Step 7 The **ssl.enabled** configuration of the ZooKeeper service in the active cluster must be the same as that in the DR cluster. Modify the **ssl.enabled** parameter in the cluster where no operation is performed by referring to the preceding steps.

Step 8 Log in to the active OMS node in the active cluster as user **root** and run the following commands to restart the DR management process:

su - omm

${BIGDATA_HOME}/om-server/om/share/om/disaster/sbin/restart-disaster.sh

If the following information is displayed, the operation is successful:

```
...
disaster start with process id : 23256
End into restart-disaster.sh
```

Step 9 Log in to the active OMS node in the DR cluster as user **root** and run the following commands to restart the DR management process:

su - omm

${BIGDATA_HOME}/om-server/om/share/om/disaster/sbin/restart-disaster.sh

Step 10 (Optional) If the cluster uses Spark services, log in to the Spark client node as user **root** and run the following commands to refresh the built-in client of the cluster:

su - omm

sh /opt/executor/bin/refresh-client-config.sh

Step 11 (Optional) If the cluster uses Flink services, log in to the node where the Flink client is installed as user **root** and run the following command to modify the Flink configuration file:

cd Client installation directory**/client/Flink/flink/conf**

vim flink-conf.yaml

Add the following parameters to the end of **env.java.opts** and save the file:

```
-Dzookeeper.clientCnxnSocket=ClientCnxnSocketNetty -Dzookeeper.client.secure=true
```


Step 12 (Optional) If the cluster uses the HetuEngine service, log in to Manager, restart HSBroker, and then log in to the HetuEngine web UI to restart all compute instances.

----End

3.4.19 Encrypting the Communication Between the Controller and the Agent

Scenario

After a cluster is installed, Controller and Agent need to communicate with each other. The Kerberos authentication is used during the communication. By default, the communication is not encrypted during the communication for the sake of cluster performance. Users who have demanding security requirements can use the method described in this section for encryption.

Impact on the System

- Controller and all Agents automatically restart, which interrupts FusionInsight Manager.
- The performance of management nodes deteriorates in large clusters. You are advised to enable the encryption function for clusters with a maximum of 200 nodes.

Prerequisites

You have obtained the IP addresses of the active and standby management nodes.

Procedure

Step 1 Log in to the active management node as user **omm**.

Step 2 Run the following command to disable logout upon timeout:

TMOUT=0

NOTE

After the operations in this section are complete, run the **TMOUT=***Timeout interval* command to restore the timeout interval in a timely manner. For example, **TMOUT=600** indicates that a user is logged out if the user does not perform any operation within 600 seconds.

Step 3 Run the following command to go to the related directory:

cd ${CONTROLLER_HOME}/sbin

Step 4 Run the following command to enable communication encryption:

./enableRPCEncrypt.sh -t

Run the **sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh** command to check whether **ResHAStatus** of the active management node Controller is **Normal** and whether you can log in to FusionInsight Manager again. If yes, the enablement is successful.

Step 5 Run the following command to disable communication encryption when necessary:

./enableRPCEncrypt.sh -f

Run the **sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh** command to check whether **ResHAStatus** of the active management node Controller is **Normal** and whether you can log in to FusionInsight Manager again. If yes, the enablement is successful.

----End

3.4.20 Updating SSH Keys for User omm

Scenario

During cluster installation, the system automatically generates the SSH public key and private key for user **omm** to establish the trust relationship between nodes. After the cluster is installed, if the original keys are accidentally disclosed or new keys are used, the system administrator can perform the following operations to manually change the keys.

Prerequisites

- The cluster has been stopped.
- No other management operations are being performed.

Procedure

Step 1 Log in as user **omm** to the node whose SSH keys need to be replaced.

If the node is a Manager management node, run the following command on the active management node.

Step 2 Run the following command to disable logout upon timeout:

TMOUT=0

NOTE

After the operations in this section are complete, run the **TMOUT=***Timeout interval* command to restore the timeout interval in a timely manner. For example, **TMOUT=600** indicates that a user is logged out if the user does not perform any operation within 600 seconds.

Step 3 Run the following command to generate a key for the node:

- If the node is a Manager management node, run the following command:

  sh ${CONTROLLER_HOME}/sbin/update-ssh-key.sh

- If the node is a non-Manager management node, run the following command:

  sh ${NODE_AGENT_HOME}/bin/update-ssh-key.sh

If "Succeed to update ssh private key." is displayed when the preceding command is executed, the SSH key is generated successfully.

Step 4 Run the following command to transfer the node's public key to the primary management node. Note that this step is necessary even if the current node is the primary management node.

scp ${HOME}/.ssh/id_rsa.pub** *oms_ip*:**${HOME}/.ssh/id_rsa.pub_bak

oms_ip: indicates the IP address of the active management node.

Enter the password of user **omm** to copy the files.

Step 5 Log in to the active management node as user **omm**.

Step 6 Run the following command to disable logout on system timeout:

TMOUT=0

Step 7 Run the following command to go to the related directory:

cd ${HOME}/.ssh

Step 8 Run the following command to add new public keys:

cat id_rsa.pub_bak >> authorized_keys

Step 9 Run the following command to move the temporary public key file, for example, **/tmp**.

mv -f id_rsa.pub_bak /tmp

Step 10 Copy the **authorized_keys** file of the active management node to the other nodes in the cluster:

scp authorized_keys *node_ip*:**/${HOME}/.ssh/authorized_keys**

node_ip: indicates the IP address of another node in the cluster. Multiple IP addresses are not supported.

Step 11 Run the following command to confirm private key replacement without entering the password:

ssh *node_ip*

node_ip: indicates the IP address of another node in the cluster. Multiple IP addresses are not supported.

Step 12 Log in to FusionInsight Manager and click **Start** in the upper right corner of **Homepage** to start the cluster.

----End

3.4.21 Changing the Access Key of an Encrypted Disk

Scenario

When a key is manually generated for disk partition encryption, the administrator needs to periodically change the access key to ensure the security of the data key.

NOTE

This section applies only to physical machine clusters.

Prerequisites

- You have obtained the SetupTool, and upload the tool script to the active management node.
- The **root** users of all nodes share the same password.

Procedure

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

Step 2 Run the following command to disable logout upon timeout:

TMOUT=0

NOTE

After the operations in this section are complete, run the **TMOUT=***Timeout interval* command to restore the timeout interval in a timely manner. For example, **TMOUT=600** indicates that a user is logged out if the user does not perform any operation within 600 seconds.

Step 3 Run the following command to go to the directory where SetupTool is decompressed, for example, the **cluster** directory in **/opt/FusionInsight_SetupTool**.

cd /opt/FusionInsight_SetupTool/preinstall/tools/cluster

Step 4 Edit the **cluster.ini** file and modify parameters as prompted.

vi cluster.ini

The following is an example:

```
g_hosts="10.xxx.xxx.xxx,10.xxx.xxx.xxx,10.xxx.xxx.xxx"
g_user_name="root"
g_password=""
g_port=22
g_timeout=500
```

Step 5 Run the following command to change the access key to a new key file.

sh replace_cluster_key.sh -k *Access key*

In this command, *Access key* indicates a specified key to be changed and its value can be **accessKey1** or **accessKey2**.

accessKey1 indicates a primary access key, and **accessKey2** indicates a secondary access key.

For example, to change the primary access key, run the following command:

sh replace_cluster_key.sh -k accessKey1

If the following information is displayed, the access key is successfully changed. By default, the primary access keys are changed for all nodes whose partitions are encrypted in the cluster.

```
Success to replace cluster key ......
```

----End

3.4.22 Hardening the OS

Scenario

Operating system security is the basis of big data system security. Configure the OS security by referring to **Table 3-14**.

Table 3-14 OS security hardening information overview

Item	Description	Risk Not Configured
Adding an OS account for O&M on each node in the cluster	The big data cluster service runs as user omm. You are advised to use an independent system user for cluster O&M.	1. If one service is compromised, other services will be affected. 2. The SSH password of user **omm** is leaked. As a result, attackers can log in to the cluster background, which damages the system.
Disabling the **history** command of the operating system on each node in the cluster	1. Log in to a node in the cluster as user **root**. 2. Run the **vi /etc/profile** command to change all HISTSIZE values in the file to **0**. 3. Run the **source /etc/profile** command for the configuration to take effect.	You can run the **history** command to view historical command records. If a user enters a password during task execution (for example, logging in to a database), sensitive information such as the password may be disclosed.

3.4.23 Enabling SSL-encrypted Transmission for the OMS Database

Scenario

By default, the data transmitted between the active and standby OMS databases is not encrypted. If the active and standby OMS databases are deployed on an untrusted network, the transmitted data may be leaked or tampered with.

This section describes how to enable Secure Sockets Layer (SSL) encrypted transmission between the active and standby OMS databases after the OMS nodes are installed.

NOTE

By default, the cluster network environment is secure, and SSL-encrypted transmission does not need to be enabled between the active and standby databases.

Prerequisites

- The OMS nodes have been installed.

Procedure

Step 1 Use PuTTY to log in to the active OMS node as user **root** and run the **su - ommdba** command to switch to user **ommdba**.

Step 2 Go to the **${GAUSSHOME}/../** directory as user **ommdba** and run the following command to enable SSL authentication between the active and standby database processes:

cd ${GAUSSHOME}/../

sh update_database_ssl.sh on

If the following information is displayed, the configuration is successful:

```
Enable database SSL successfully.
```

Step 3 Log in to the standby OMS node as user **ommdba** and perform **Step 2**.

----End

3.4.24 Disabling SSL-encrypted Transmission for the OMS Database

Scenario

By default, the data transmitted between the active and standby OMS databases is not encrypted. If the active and standby OMS databases are deployed on an untrusted network, the transmitted data may be leaked or tampered with.

This section describes how to disable Secure Sockets Layer (SSL) encrypted transmission between the active and standby OMS databases after the OMS nodes are installed.

NOTE

By default, the cluster network environment is secure, and SSL-encrypted transmission does not need to be enabled between the active and standby databases.

Prerequisites

- The cluster has been installed.
- SSL-encrypted transmission between the active and standby OMS databases has been enabled.

Procedure

Step 1 Use PuTTY to log in to the active OMS node as user **root** and run the **su - ommdba** command to switch to user **ommdba**.

Step 2 Go to the **${GAUSSHOME}/../** directory and run the following command to disable SSL authentication between the active and standby database processes:

sh update_database_ssl.sh off

If the following information is displayed, the configuration is successful:

```
Disable database SSL successfully.
```

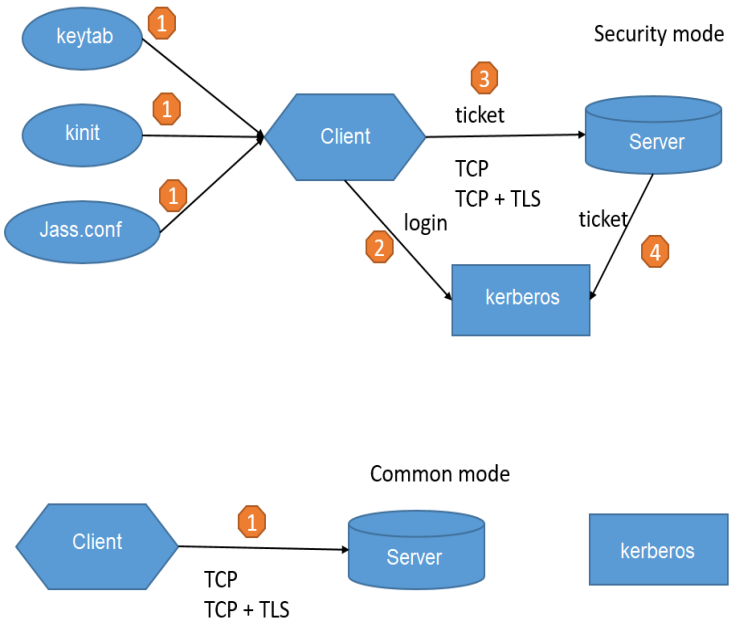
Step 3 Log in to the standby OMS node as user **ommdba** and perform **Step 2**.

----End

3.4.25 Hardening the Security of Redis

Authentication and Encryption

The authentication system of Redis is as follows:



- The common mode does not require authentication. You can use the built-in default user to log in to the system.
- In security mode, clients can be authenticated in keytab(Jedis), kinit(redis-cli), or Jaas.conf (Jedis) mode.
- Redis permissions:

Resource	Permission
Read permission	Only Redis read commands, such as GET and HGET, can be executed.

Resource	Permission
Write permission	Redis write commands, such as SET and ZADD, can be executed.
Management permission	Redis management commands, such as SAVE and BGSAVE, can be executed.

NOTE

- Currently, Kerberos authentication cannot be enabled for Redis in non-security mode.
- After the suppergroup permission is granted to a common user, the user has the read, write, and management permissions on Redis by default. You do not need to set the permissions separately.
- User **admin** is not allowed to access Redis service data.

Channel Encryption

Redis supports open source channel encryption, ensuring data security during transmission.

Security Hardening

- Removing insecure passwords

  Note: Kerberos authentication and Manager-based permission management are supported.

  Security hardening: In security mode, Redis supports only Kerberos authentication and does not support the username and password authentication mode. In addition, Redis 6 does not support the ACL feature. Therefore, passwords are not encrypted and saved to the local host.

- Disabling high-risk commands

  Note: High-risk commands are disabled for Redis.

  Security hardening: To prevent irreversible impact on Redis caused by high-risk commands, **FLUSHALL**, **FLUSHDB**, **EVAL**, **KEYS**, **SAVE**, and **DEBUG** are disabled by default.

- Anti-repudiation

  Description: Audit logs are recorded.

  Security hardening: Login users and risky operations are audited.

3.4.26 Solr Security Hardening

Anonymizing Return Values of Solr Interfaces

Scenario

Some Solr interfaces display sensitive information such as JVM environment variables, data directories, and version numbers. When an interface is abnormal, detailed stack information is displayed, which poses security risks.

Enable Solr security hardening to anonymize the data returned by Solr interfaces for higher security.

Security Hardening Points

- Anonymize the versions and JVM information displayed on the **Dashboard** tab page of the Solr web UI.
- Anonymize the return value of the /admin/system interface of Solr.
- Do not return detailed stack information if an exception occurs when Solr calls an interface.

Procedure

Security hardening is disabled for Solr by default. You can configure the following parameters to enable security hardening:

Step 1 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **Solr**. On the page that is displayed, click the **Configurations** tab then the **All Configurations** sub-tab.

Step 2 Search for the following parameter in the search box and change its value to **true**:

Table 3-15 Solr security hardening parameter

Parameter	Description	Default Value
SOLR_REDACTION_SENSITIVE_ENABLED	Whether to enable Solr security hardening. Once enabled, the returned values of commands will be anonymized.	false

Step 3 After the modification is complete, click **Save** then **OK**.

Step 4 On the **Dashboard** tab page of the Solr service, choose **More** > **Restart Service**. In the **Restart Service** dialog box that is displayed, enter the password for verification and click **OK**.

----End

Fixing the Solr 0-day Vulnerability

Scenario

Solr clusters in normal mode will be attacked and any files can be read if attackers enable **requestDispatcher.requestParsers.enableRemoteStreaming** using the Config API.

Procedure

1. The streaming function of Solr is disabled by default. Attackers need to use the Config API to enable this function. To fix the vulnerability, the public parameter **requestDispatcher.requestParsers.enableRemoteStreaming** is disabled to enable the streaming function. When this parameter is used to enable this function, the following exception is thrown:

```
{
  "responseHeader":{
    "status":400,
    "QTime":1},
  "errorMessages":["error processing commands\n"],
```



```
 "WARNING":"This response format is experimental.  It is likely to change in the future.",
 "error":{
   "details":[{
       "set-property":{"requestDispatcher.requestParsers.enableRemoteStreaming":true},
       "errorMessages":["Enabling the RemoteStreaming function may expose sensitive server
information, which poses security risks."]}],
   "msg":"error processing commands",
   "code":400}}
```

2. The streaming function can be forcibly enabled by configuring parameters. The following information indicates that the streaming function is enabled.

```
{
 "responseHeader":{
   "status":0,
   "QTime":486},
 "WARNING":"This response format is experimental.  It is likely to change in the future."}
```

3.4.27 Configuring FTP-Server for Encrypted Data Transmission

Scenario

Configure FTPS for data transmission between nodes in a cluster for higher security.

NOTE

You are advised to use FTPS. If an insecure protocol is used, risks may occur.

Procedure

Step 1 Log in to FusionInsight Manager and choose **Cluster** > **Services** > **FTP-Server**. Click **Configurations** then **All Configurations** and enter the parameter name in in the search box.

Table 3-16 Security configuration items

Item	Description	New Value
ftp-enabled	Whether to enable FTP. You are advised to use FTPS on a non-isolated network. The default value is **false**.	false
ftps-enabled	Whether to enable FTPS. The default value is **true**.	true **NOTE** If this parameter is set to **false**, FTP is used. Ensure that the service works in a secure network environment.

Step 2 Click **Save** to save the configuration.

Step 3 Click **Dashboard** to go to the FTP-Server service page. Choose **More** > **Restart Service**. Verify the identity and click **OK**. Wait until the restart is successful.

----End

3.5 Security Maintenance

3.5.1 Managing the Rights of User omm

Scenario

You can use the OS user **omm** of the node install clusters. When installing clusters as the OS user **omm** of the node, you need to modify the rights of user **omm** on all involved nodes in node scaling scenarios.

NOTE

This section applies only to physical machine clusters.

Prerequisites

- You have obtained the account of user **root** of each node in the cluster. In a node scaling-out scenario, the accounts of user **root** of the new nodes are also required. User **root** on each node has the remote login rights.
- You have decompressed and saved the required software package on the active management node.

Granting Operation Rights to User omm

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

Step 2 Copy the **/opt/FusionInsight_SetupTool/preset** folder on the active management node to the **/opt** directories of all nodes and run the **sh /opt/preset/preset.sh** command on these nodes.

You can use the script tool in the software package to run a command or access a file on multiple nodes in a cluster. For details, see **How Do I Run Commands or Access Files on Multiple Nodes in a Cluster?**.

Step 3 Run the following command as user **root** on active and standby management nodes to set the owner of the software package directory. **/opt/FusionInsight_Manager/** indicates the directory generated after the software package is decompressed.

chown omm: /opt/FusionInsight_Manager/ -R

----End

Deleting Rights of User omm

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

Step 2 Run the **sh /opt/preset/postset.sh** command on all nodes.

----End

3.5.2 Account Maintenance Suggestions

It is recommended that the administrator conduct routine checks on the accounts. The check covers the following items:

- Check whether the accounts of the OS, FusionInsight Manager, and each component are necessary and whether temporary accounts have been deleted.
- Check whether the permissions of the accounts are appropriate. Different administrators have different rights.
- Check and audit the logins and operation records of all types of accounts.

3.5.3 Password Maintenance Suggestions

Accessing portal requires identity authentication. The complexity and validity period of an account password must meet your security requirements.

Refer to the following suggestions to maintain passwords:

1. Assign dedicated personnel to keep OS passwords.
2. Use passwords that meet certain strength requirements, such as minimum password length or mixing of letter cases.
3. Encrypt passwords before transferring them, and do not transfer them via email.
4. Encrypt passwords for storage.
5. Remind enterprise users to change passwords during system handover.
6. Change passwords periodically.

3.5.4 Log Maintenance Suggestions

Operation logs help discover exceptions such as illegal operations and login by unauthorized users. The system records important operations in logs. You can use operation logs to locate problems.

Checking Logs Regularly

Check system logs periodically and handle exceptions such as unauthorized operations or logins in a timely manner.

Backing Up Logs Regularly

The audit logs provided by FusionInsight Manager and cluster record the user activities and operations. You can export the audit logs on FusionInsight Manager. If there are too many audit logs in the system, you can configure dump parameters to dump audit logs to a specified server to ensure that the cluster nodes disk space is sufficient.

Maintenance Owner

Network monitoring engineers and system maintenance engineers

3.5.5 OS Maintenance Suggestions

Check and harden OS security according to **Table 3-17**; otherwise, system cannot be used properly after configuration items are modified.

Table 3-17 OS security hardening information overview

Check Item	Association Relationship	Operation Suggestion
Check the files that have the **suid** and **sgid** permissions on each node.	The **/bin/ping** command is used in the system.	It is recommended that the command permission remain the same on each node. **NOTE** By default, the **sgid** folder is created in the Hadoop data directory. In this directory, temporary files generated during YARN job running must have the same owner group as the **sgid** directory.
Check whether the **root** user remote login function is disabled on each node.	The FusionInsight installation, uninstallation, and capacity expansion require the root user remote login permission.	Enable the remote login by user **root** only when the **root** user is used to perform installation, uninstallation, expansion, or restoration operations. Run the following command on the cluster deployment node: **sed -i "s/PermitRootLogin yes/PermitRootLogin no/g" /etc/ssh/sshd_config** **service sshd restart**
Check the permission of key directories or files on each node.	The system internal running user **omm** requires the **/etc** directory access permission.	It is recommended that the **/etc** directory permission remain the same on each node.
Check whether the interactive system account login function is disabled on each node.	Nodes in the cluster need interactive login of users **omm** and **root**.	It is recommended that the interactive login permission be retained for **omm** and **root** nodes in the cluster. Do not use the **omm** user to remotely access the OS.

Check Item	Association Relationship	Operation Suggestion
Check whether the NTP service is running on each node.	System depends on the NTP service.	It is recommended that the NTP service be running properly on each node. The involved services are as follows: • ntpd • ntpdate
Check whether the services that OpenLDAP depends on are running on each node.	The system uses OpenLDAP, which depends on specific services.	It is recommended that the services that OpenLDAP depends on be running properly on each node. The involved services are as follows: • nscd/sssd • slapd
Check whether common OS users are allowed to perform cron scheduled tasks on each node.	The running user **omm** of the system will perform cron scheduled tasks.	It is recommended that user **omm** in the OS be allowed to perform cron scheduled tasks on each node.
Check whether OS logs, such as the OS logs in the **/var/log/messages** and **/var/log/secure** directories, are archived and compressed periodically on each node.	Logs will be accumulated if they are not archived periodically until the disk space is insufficient. As a result, system will fail to record logs. **NOTE** The OpenLDAP run log function is disabled by default to prevent a large number of OS logs being generated to affect the proper running of the Syslog service.	It is recommended that the logrotate service or the cron service of the OS are used to archive and compress OS logs, and delete expired OS logs periodically on each node.

Check Item	Association Relationship	Operation Suggestion
Check whether the password lock mechanism is set for the **omm** and **ommdba** accounts created on FusionInsight Manager.	If the password lock mechanism is not set, all OS accounts (including the **omm** and **ommdba** accounts) will be under security risks.	It is recommended that you set the password lock mechanism for all OS accounts on each node. • Modify the **/etc/pam.d/system-auth** file to set the password lock mechanism for the Red Hat OS accounts. • Modify the **/etc/pam.d/login** and **/etc/pam.d/sshd** files to set the password lock mechanism for the SUSE OS accounts.
Check whether the challenge response threshold is proper.	If the Linux kernel version is between v3.6 and v4.7, the challenge response in the TCP implementation of the kernel may cause information leakage. Therefore, the value of **tcp_challenge_ack_limit** needs to be adjusted to a large range so that attackers cannot reach it properly and no additional data of the client server connection can be inferred.	For example, if the default value of **tcp_challenge_ack_limit** is **1000**. You can change it to **999999999**. **NOTE** For EulerOS, this parameter is automatically adjusted during the preinstall process. For other OSs, you are advised to set this parameter as required.

3.5.6 Security Emergency Response Mechanism

Users are required to set up an emergency plan as a quick response to security accidents. This helps recover production, resolve problems, and minimize loss.

If your system is faulty, try to rectify the fault by referring to the documents delivered together with devices. The documentation provides guidance on how to solve common problems that occur during routine maintenance and troubleshooting.

If the document do not help and fault persists, contact technical support engineers for assistance. You are advised to collect necessary fault information and make debugging preparations before contacting Huawei technical support engineers.

3.5.7 Emergency Response Email Address

When you encounter the following situations, contact Huawei PSIRT at psirt@huawei.com:

1. Provide feedback on vulnerabilities of Huawei products.
2. Obtain security emergency response services from Huawei.
3. Obtain information about vulnerabilities of Huawei products.

Encrypt sensitive information before sending. Obtain a cipher key from:

http://www.huawei.com/en/security/psirt/about-huawei-psirt/index.htm

3.6 Security Statements

The user security information is involved when MRS is used. Before using MRS, observe the following:

- The Virtual Private Cloud (VPC) and security group (SG) are used to protect cluster data. The management plane can access the cluster using SSH only during cluster creation, scale-ins, and scale-outs. In other cases, the management plane can only invoke REST APIs through tokens to access cluster services.
- Hadoop open-source APIs do not provide independent authentication. When submitting jobs, ensure that the jobs do not contain viruses or cause unrecoverable destructive operations on the cluster.
- Job run logs can be stored in the OBS or HDFS directory specified by users.
- Clusters and jobs cannot be restored after being deleted. If a cluster is deleted before data analysis or processing is complete, the data may be lost. Therefore, exercise caution when performing this operation.

JDK Usage Statement

An MRS cluster is a big data cluster that provides users with distributed data analysis and computing capabilities. The built-in JDK of the product is OpenJDK, which is used in the following scenarios:

- Platform service running and maintenance
- Linux client operations, including service submission and application O&M

JDK Risk Description

The system performs permission control on the built-in JDK. Only users in the related group of the FusionInsight platform can access the JDK. In addition, the platform is deployed on the customer's intranet. Therefore, the security risk is low.

JDK Hardening

For details, see "Hardening JDK" in **Hardening Policies**.

Public IP Addresses in Hue

Hue uses the test cases of third-party packages, such as **ipadrress**, **requests**, and **Django**, and uses the public IP addresses in the comments of the test cases. However, these public IP addresses are not involved when Hue provides services, and the Hue configuration file does not involve these public IP addresses.

3.7 Appendix

3.7.1 FAQ

3.7.1.1 Logging In to FusionInsight Manager

Scenario

Log in to FusionInsight Manager using an account.

Procedure

Step 1 Obtain the URL for logging in to FusionInsight Manager.

Step 2 On login page, enter the username and password.

The default username is **admin**. To obtain the default password, see ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator.

Step 3 Change the password upon your first login.

The password must meet the following complexity requirements:

- Contains 8 to 64 characters.
- Contains at least four types of the following: uppercase letters, lowercase letters, numbers, spaces, and special characters (`~!@#$%^&*()-_=+|[{}];',<.>/\?).
- Cannot be the same as the username or the username spelled backwards.
- Cannot be the same as the current password.

Step 4 Move the cursor over  in the upper right corner of home page, and choose **Logout** from the drop-down list. In the dialog box that is displayed, click **OK** to log out of the current user.

----End

3.7.1.2 Logging In to the Management Node

Scenario

Some O&M operation scripts and commands need to be run or can be run only on the active management node. You can identify and log in to the active or standby management node based on the following operations.

Checking and Logging In to the Active and Standby Management Nodes

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **OMS**.

In the **Basic Information** area, **Current Active** indicates the host name of the active management node, and **Current Standby** indicates the host name of the standby management node.

Click a host name to go to the host details page. On the host details page, record the IP address of the host.

Step 3 Log in to the active or standby management node as user **root**.

----End

Identifying the Active and Standby Management Nodes by Running Scripts and Logging In to Them

Step 1 Log in to any node where FusionInsight Manager is installed as user **root**.

Step 2 Run the following command to identify the active and standby management nodes:

su - omm

sh ${BIGDATA_HOME}/om-server/om/sbin/status-oms.sh

In the command output, the node whose **HAActive** is **active** is the active management node (Master1), and the node whose **HAActive** is **standby** is the standby node (Master2).

```
HAMode
double
NodeName            HostName      HAVersion       StartTime            HAActive
HAAllResOK          HARunPhase
192-168-0-30        Master1       V100R001C01     2021-09-01 07:12:05  active
normal              Actived
192-168-0-24        Master2       V100R001C01     2021-09-01 07:14:02  standby
normal              Deactived
```

Step 3 Run the following command to obtain the IP addresses of the active and standby management nodes:

cat /etc/hosts

Example IP addresses of the active and standby management nodes:

```
127.0.0.1       localhost
192.168.0.30    Master1
192.168.0.24    Master2
```

Step 4 Log in to the active or standby management node as user **root**.

----End

3.7.1.3 Configuring Password Policies

Scenario

To keep up with service security requirements, you can set password security rules, user login security rules, and user locking rules on FusionInsight Manager.

NOTICE

- Modify password policies based on service security requirements, because they involve user management security. Otherwise, security risks may be incurred.
- Change the user password after modifying the password policy, and then the new password policy can take effect.
- This password policy applies to human-machine accounts created on Manager.

Adding a Password Policy

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **Security Policy** > **Password Policy**.

Step 3 Click **Add Password Policy** and modify the password policy as prompted.

For details about the parameters, see **Table 3-18**.

Table 3-18 Password policy parameters

Parameter	Description
Password Policy Name	The value is a string of 3 to 32 characters, including case-insensitive letters, digits, underscores (_), and hyphens (-). It cannot start with a hyphen (-).
Minimum Password Length	Indicates the minimum number of characters a password contains. The value ranges from **8** to **32**.
Character Types	Indicates how many character types in the following types a password can contain at least: uppercase letters, lowercase letters, digits, spaces, and special characters (~`!?,.:;-_'(){}[]/<>@#$%^&*+|\=). The value can be **4** or **5**. The default value is **4**, which means that a password can contain uppercase letters, lowercase letters, digits, and special characters. If you set the parameter to **5**, a password can contain all the five character types mentioned above.
Password Retries	Indicates the number of consecutive wrong password attempts allowed before the system locks the user. The value ranges from **3** to **30**.
User Lock Duration (Min)	Indicates the time period in which a user is locked when the user lockout conditions are met. The value ranges from **5** to **120**.
Password Validity Period (Day)	Indicates the validity period of a password. The value ranges from **0** to **90**. **0** indicates that the password is permanently valid.

Parameter	Description
Repetition Rule	Indicates the number of previous passwords that cannot be reused when you change the password. The value ranges from **1** to **5**. The default value is **1**. This policy applies to only human-machine accounts.
Password Expiration Notification (Days)	Indicates the number of days in advance users are notified that their passwords are about to expire. After the value is set, if the difference between the cluster time and the password expiration time is smaller than this value, the user receives password expiration notifications. When logging in to FusionInsight Manager, the user will be notified that the password is about to expire and a message is displayed asking the user to change the password. The value ranges from **0** to *X* (*X* must be set to the half of the password validity period and rounded down). Value **0** indicates that no notification is sent.
Interval for Deleting Authentication Failure Records (Min)	Indicates the interval of retaining incorrect password attempts. The value ranges from **0** to **1440**. **0** indicates that incorrect password attempts are permanently retained, and **1440** indicates that incorrect password attempts are retained for one day.

Step 4 Click **OK** to save the configurations.

A new user uses the default password policy. After a new password policy is created, you can manually select the password policy when creating a user. You can modify the password policy of an existing user. For details, see **Modifying User Information**.

----End

NOTE

A maximum of 32 password policies can be created.

Modifying a Password Policy

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **Security Policy** > **Password Policy**.

Step 3 Click **Modify** in the row that contains the target password policy. On the **Modify Password Policy** page, modify the password policy as prompted.

For details about the parameters, see **Table 3-18**.

Step 4 Click **OK** to save the configurations.

----End

NOTE

- Users (except **admin**) cannot modify their own password policies.
- After the password policy bound to a user is modified, if the remaining password validity period is greater than the password validity period in the new password policy, the password validity period is set to the validity period in the new password policy. If the remaining password validity period is less than the password validity period in the new password policy, the password validity period remains unchanged.

Deleting a Password Policy

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **Security Policy** > **Password Policy**.

Step 3 Click **Delete** in the row that contains the target password policy. In the dialog box that is displayed, click **OK**.

----End

NOTE

The default password policy and the password policy that has been bound to a user cannot be deleted.

3.7.1.4 User Management

3.7.1.4.1 Creating a User

Scenario

FusionInsight Manager supports a maximum of 50,000 users (including built-in users). By default, only user **admin** has the highest operation permissions of FusionInsight Manager. You need to create users on FusionInsight Manager and assign operation permissions to the users based on service requirements.

NOTE

Information about the newly created user is synchronized to the OS caches of all nodes in the cluster. The value of **uid** of the new user ranges from **20000** to **100000**. You can run the **id** *Username* command on the node to check the value.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 On the **User** page, click **Create**.

Step 4 Set **Username**. The username can contain digits, letters, underscores (_), hyphens (-), and spaces. It is case-insensitive and cannot be the same as any existing username in the system or OS.

Step 5 Set **User Type** to **Human-Machine** or **Machine-Machine**.

- **Human-Machine** user: used for FusionInsight Manager O&M and component client operations. If you select this option, you also need to select the password policy and set **Password** and **Confirm Password**.
- **Machine-Machine** user: used for component application development. If you select this option, the password is randomly generated.

Step 6 In the **User Group** area, click **Add** to add one or more user groups to the list.

NOTE

- If the selected user group has been bound to a role or a permission policy has been configured in Ranger, the user can obtain the corresponding permissions.
- After FusionInsight Manager is installed, some user groups generated by default have special permissions. Select desired user groups based on the descriptions on the UI.
- If existing user groups cannot meet your requirements, click **Create User Group** to create a user group. For details, see **Creating a User Group**.

Step 7 Select a group from the **Primary Group** drop-down list to create directories and files.

The drop-down list contains all groups selected in **User Group**.

NOTE

A user can belong to multiple groups (including the primary group and secondary groups). The primary group is set to facilitate maintenance and comply with the permission mechanism of the Hadoop community. The primary group has the same permission control functionality as other groups.

Step 8 In the **Role** area, click **Add** to bind roles to the user.

NOTE

- Adding a role when you create a user can specify the user permissions.
- If the permissions granted to the user from the user group cannot meet service requirements, you can bind other created roles to the user. You can also click **Create Role** to create a role first. For details, see **Creating a Role**.

  It takes 3 minutes to make role permission assignment to the user take effect. If the permissions obtained from the user group are enough, you do not need to add a role.
- After Ranger authentication is enabled for a component, you need to configure Ranger policies to assign permissions to the user except the permissions of default user group or role.
- If a user is not added to a user group or assigned a role, the user cannot view information or perform operations after logging in to FusionInsight Manager.

Step 9 Enter information in **Description**.

Step 10 Click **OK**.

After a human-machine user is created, you need to change the initial password as prompted after logging in to FusionInsight Manager.

If storage-compute decoupling is used in the current cluster and the new user has the permission to delete resources related to components interconnected with OBS 3.0, you need to configure lifecycle management policies for the recycle bin directory of the user in the OBS 3.0 file system. For details, see "Configuring the Policy for Clearing Component Data in the Recycle Bin" in *MapReduce Service*

(MRS) 3.3.1-LTS User Guide (for Huawei Cloud Stack 8.3.1) in the MapReduce Service (MRS) 3.3.1-LTS Usage Guide (for Huawei Cloud Stack 8.3.1).

----End

3.7.1.4.2 Modifying User Information

Scenario

You can modify user information on FusionInsight Manager, including the user group, primary group, role permission assignment, and user description.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 Locate the row that contains the target user and click **Modify** in the **Operation** column.

Modify the parameters based on service requirements.

NOTE

- It takes three minutes at most for the change of the user group or role permissions to take effect.
- Users (except **admin**) cannot modify their own password policies.
- Locked users cannot modify their password policies.
- After the password policy bound to a user is modified, the modification takes effect when the user changes the password next time.
- After the password policy bound to a user is modified, if the remaining password validity period is greater than the password validity period in the new password policy, the password validity period is set to the validity period in the new password policy. If the remaining password validity period is less than the password validity period in the new password policy, the password validity period remains unchanged.

Step 4 Click **OK**.

----End

3.7.1.4.3 Changing a User Password

Scenario

For security purposes, the password of a human-machine user must be changed periodically.

If users have the permission to use FusionInsight Manager, they can change their passwords on FusionInsight Manager.

If users do not have the permission to use FusionInsight Manager, they can change their passwords on the client.

Prerequisites

- You have obtained the current password policy.
- The user has installed the client on any node in the cluster and obtained the IP address of the node. The password of the client installation user can be obtained from the administrator.

Changing the Password on FusionInsight Manager

Step 1 Log in to FusionInsight Manager.

Step 2 Move the cursor to the username in the upper right corner of the page.

On the user account drop-down menu, choose **Change Password**.

Step 3 On the displayed page, set **Current Password**, **New Password**, and **Confirm Password**, and click **OK**.

By default, the password must meet the following complexity requirements:

- The password contains at least 8 characters.
- The password must contain at least four types of the following characters: uppercase letters, lowercase letters, digits, spaces, and special characters (`~!@#$%^&*()-_=+|[{}];',<.>/\?).
- The password cannot be the same as the username or the username spelled backwards.
- The password cannot be a common easily-cracked password.
- The password cannot be the same as the password used in the latest *N* times. *N* indicates the value of **Repetition Rule** configured in **Configuring Password Policies**.

----End

Changing the Password on the Client

Step 1 Log in to the node where the client is installed as the client installation user.

Step 2 Run the following command to switch to the client directory, for example, **/opt/hadoopclient**:

cd /opt/hadoopclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

Step 4 Change the user password. This operation takes effect for all servers.

kpasswd *System username*

For example, if you want to change the password of system user **test1**, run the **kpasswd test1** command.

By default, the password must meet the following complexity requirements:

- The password contains at least 8 characters.

- The password must contain at least four types of the following characters: uppercase letters, lowercase letters, digits, spaces, and special characters (`~!@#$%^&*()-_=+|[{}];',<.>/\?).
- The password cannot be the same as the username or the username spelled backwards.
- The password cannot be a common easily-cracked password.
- The password cannot be the same as the password used in the latest *N* times. *N* indicates the value of **Repetition Rule** configured in **Configuring Password Policies**.

NOTE

If an error occurs during the running of the **kpasswd** command, try the following operations:

- Stop the SSH session and start it again.
- Run the **kdestroy** command and then run the **kpasswd** command again.

----End

3.7.1.4.4 Managing User Groups

Scenario

FusionInsight Manager supports a maximum of 5000 user groups (including built-in user groups). You can create and manage different user groups based on service scenarios on FusionInsight Manager. A user group is bound to a role to obtain operation permissions. After a user is added to a user group, the user can obtain the operation permissions of the user group. A user group can be used to classify users and manage multiple users.

NOTE

Information about the newly created user group is synchronized to the OS cache of all nodes in the cluster. The value range of **gid** for the newly created user group is 8000–8999, 9998, and 10000–300000.

Prerequisites

- You have learned service requirements and created roles required by service scenarios.
- You have logged in to FusionInsight Manager.

Creating a User Group

Step 1 Choose **System** > **Permission** > **User Group**.

Step 2 Above the user group list, click **Create User Group**.

Step 3 Set **Group Name** and **Description**.

The group name contains 1 to 64 characters, including case-insensitive letters, digits, underscores (_), hyphens (-), and spaces. It cannot be the same as an existing user group name in the system.

Step 4 In the **Role** area, click **Add** to select a role and add it.

NOTE

- For components (except HDFS and Yarn) for which Ranger authorization has been enabled, the permissions of non-default roles on Manager do not take effect. You need to configure Ranger policies to assign permissions to user groups.
- If the resource requests of HDFS and Yarn are beyond the Ranger policies, the ACL rules of the components still take effect.

Step 5 In the **User** area, click **Add** to select a user and add it.

Step 6 Click **OK**.

The user group is created.

----End

Viewing User Group Information

By default, all user groups are displayed in the user group list. You can click the arrow on the left of a user group name to view details about the user group, including the user quantity, specific users, and bound roles of the user group.

Modifying Information About a User Group

Locate the row that contains the target user group, and click **Modify** to modify its information.

Exporting Information About a User Group

Click **Export All** to export all user group information at a time in **TXT** or **CSV** format.

The user group information contains the following fields: user group name, description, number of users, cluster, service, user list, and role list.

Deleting a User Group

Locate the row that contains the target user group, and click **Delete**. To delete multiple user groups in batches, select the target user groups and click **Delete** above the user group list. A user group that contains users cannot be deleted. To delete such a user group, delete all its users by modifying the user group first.

3.7.1.4.5 Managing Roles

Scenario

FusionInsight Manager supports a maximum of 5000 roles (including system built-in roles but excluding roles automatically created by tenants). Based on different service requirements, you need to create and manage different roles on FusionInsight Manager and perform authorization management for FusionInsight Manager and components using roles.

Prerequisites

- You have learned service requirements.

- You have logged in to FusionInsight Manager.

Creating a Role

Step 1 Choose **System** > **Permission** > **Role**.

Step 2 On the displayed page, click **Create Role** and specify **Role Name** and **Description**.

The role name consists of 3 to 50 characters, including digits, letters, and underscores (_). It cannot be the same as an existing role name in the system. The role name cannot start with **Manager**, **System**, or **default**. For example, the role name cannot be **Manager_test**.

Step 3 In the **Configure Resource Permission** area, click the cluster whose permissions are to be added and select service permissions for the role.

When setting permissions for a component, enter a resource name in the search text box in the upper right corner and click the search icon to view the search result.

The search result contains only directories, but not subdirectories. Search by keyword supports fuzzy match and is case-insensitive.

NOTE

- For components (except HDFS and Yarn) for which Ranger authorization has been enabled, the permissions of non-default roles on Manager do not take effect. You need to configure Ranger policies to assign permissions to user groups.
- If the resource requests of HDFS and Yarn are beyond the Ranger policies, the ACL rules of the components still take effect.
- A maximum of 1000 permissions can be set for a component at a time.

Step 4 Click **OK**.

----End

Modifying Role Information

Locate the row that contains the target role and click **Modify**.

Exporting Role Information

Click **Export All** to export all role information at a time in **TXT** or **CSV** format.

The role information contains the following fields: role name, description, creation time, user, and user group.

Deleting a Role

Locate the row that contains the target role and click **Delete**. To delete multiple roles in batches, select the target roles and click **Delete** above the role list. A role bound to a user cannot be deleted. To delete such a role, disassociate the role from the user by modifying the user first.

Task Example (Creating a Manager Role)

Step 1 Choose **System** > **Permission** > **Role**.

Step 2 On the displayed page, click **Create Role** and fill in **Role Name** and **Description**.

Step 3 In the **Configure Resource Permission** area, click **Manager** and set permissions for the role.

Manager permissions:

- Cluster
  - **view** permission: permission to view information on the **Cluster** page and view alarms and events under **O&M** > **Alarm**.
  - **management** permission: permission for management on the **Cluster** and **O&M** pages.
- User
  - **view** permission: permission to view information on pages under **System** > **Permission**.
  - **management** permission: permission for management on pages under **System** > **Permission**.
- Audit

  management permission: permission for management on the **Audit** page.
- Tenant

  management permission: permission for management on the **Tenant** page and permission to view alarms and events under **O&M** > **Alarm**.
- System

  management permission: permission for management on all pages except those under **Permission** on the **System** page and permission to view alarms and events under **O&M** > **Alarm**.
- Disaster

  management permission: permission to call DR APIs to connect to the peer cluster in the active/standby DR scenario.

Step 4 Click **OK**.

----End

3.7.1.4.6 Exporting an Authentication Credential File

Scenario

If a user uses a security mode cluster to develop applications, the keytab file of the user needs to be obtained for security authentication. You can export keytab files on FusionInsight Manager.

NOTE

After a user password is changed, the exported keytab file becomes invalid, and you need to export a keytab file again.

Prerequisites

Before downloading the keytab file of a Human-Machine user, the password of the user must be changed at least once on the Manager portal or a client; otherwise, the downloaded keytab file cannot be used For details, see **Changing a User Password**.

Procedure

Step 1 Log in to FusionInsight Manager.

Step 2 Choose **System** > **Permission** > **User**.

Step 3 Locate the row that contains the user whose keytab file needs to be exported, choose **More** > **Download Authentication Credential**, specify the save path after the file is automatically generated, and keep the file properly.

The authentication credential includes the **krb5.conf** file of the Kerberos service.

After the authentication credential file is decompressed, you can obtain the following two files:

- The **krb5.conf** file contains the authentication service connection information.
- The **user.keytab** file contains user authentication information.

----End

3.7.1.5 How Do I Solve the Error That Occurs During the kpasswd Command Execution?

Question

The following error occurs when the **kpasswd** command is running. How do I solve it?

```
kpasswd:Principal information error while getting initial ticket
```

Answer

- Stop the SSH session and start it again.
- Run the **kdestroy** command and then run the **kpasswd** command again.

3.7.1.6 How Do I Run Commands or Access Files on Multiple Nodes in a Cluster?

NOTE

This section applies only to physical machine clusters.

Question

During cluster installation or routine maintenance, you can use the script tool in the software package to run a command or access a file on multiple nodes in a cluster.

Answer

Prerequisites

- You have obtained the operator username and password of each node in the cluster and have enabled the user's remote login permission.
- You have decompressed the **FusionInsight_SetupTool_XXX.tar.gz** package, the script tool package used for FusionInsight software installation, to the **/opt** directory of the active management node.

Procedure

Step 1 Log in to the active management node as an operator.

Step 2 Go to the **/opt/FusionInsight_SetupTool/preinstall/tools/cluster** directory and edit the **cluster.ini** file as required.

Table 3-19 Parameters in **cluster.ini**

Parameter	Example Value	Description
g_hosts (mandatory)	192.168.10.[10-20]	Specifies the IP addresses of all nodes where the operation is performed. ● Use commas (,) to separate IP addresses. For example, **192.168.10.10**, **192.168.10.11**. ● Use hyphens (-) to indicate an IP address segment if the IP addresses are consecutive. For example, **192.168.10.[10-20]**. ● Use hyphens (-) to indicate an IP address segment if the IP addresses are consecutive, and use commas (,) to separate IP address segments. For example, **192.168.10.[10-20,30-40]**.
g_user_name	root	Specifies the user who performs the operation.

Parameter	Example Value	Description
g_password	N/A	Specifies the password file corresponding to the user who performs the operation. The password is empty by default.
g_port	22	Specifies the SSH connection port. The default value is **22**.
g_timeout	10	Specifies the SSH connection timeout period. The default value is **10 seconds**. The value increases when the network conditions are poor.
g_ip_model	IPv4	IP address mode of the network where the cluster is located.

By default, **g_password** is left blank. If the passwords of all users who perform the operation are the same, you only need to enter the password once after the command is executed. Otherwise, you must manually create a password file and set **g_password** to the full path of the password file. The password file is in the following format:

```
IP address 1
Password of IP address 1
IP address 2
Password of IP address 2
...
```

NOTE

- Node passwords stored locally have security risks. Therefore, it is recommended that the password of the executor for each node be the same. If you have to use these passwords, ensure that only executors have read and write permissions for the password file and delete the password file immediately after use.
- The password file must be in the UNIX format.
- The last line of the password file cannot be empty.
- Special characters contained in the password do not need to be converted.

For example, the content for creating the **secret.txt** file is as follows:

```
10.10.37.[10-11]
123456!654321
10.10.37.12
123456!
```

Step 3 Run the command based on the scenarios.

1. This command is executed on each node.

   Command format: **./clustercmd.sh** *Detailed command*

Example (Running the **hostname** command on each node):

```
dc-rack1007-4m:/cluster # ./clustercmd.sh hostname
==>>10.10.37.10
dc-rack1007-1
==>>10.10.37.11
dc-rack1007-2
==>>10.10.37.12==
dc-rack1007-3
```

2. Copy the file from each node to the specified directory on the node.

   Command format: **./clusterscp.sh get** *Source path Target path*

   Example (Copying the **/opt/test/mem.txt** file from each node to the **/opt/result** directory on the node):

   ```
   dc-rack1007-4m:/cluster # ./clusterscp.sh get /opt/test/mem.txt /opt/result
   get /opt/result/10.10.37.10_mem.txt from 10.10.37.10:/opt/test/mem.txt successfully.
   get /opt/result/10.10.37.11_mem.txt from 10.10.37.11:/opt/test/mem.txt successfully.
   get /opt/result/10.10.37.12_mem.txt from 10.10.37.12:/opt/test/mem.txt successfully.
   ```

3. Copy the specified files or folders from the node to the specified directory on each node.

   Command format: **./clusterscp.sh put** *Source path Target path*

   Example (Copying the **/opt/test/hosts** file from the node to the **/etc** directory on each node):

   ```
   dc-rack1007-4m:/opt/cluster # ./clusterscp.sh put /opt/test/hosts /etc
   put /opt/test/hosts to 10.10.37.10:/etc successfully.
   put /opt/test/hosts to 10.10.37.11:/etc successfully.
   put /opt/test/hosts to 10.10.37.12:/etc successfully.
   ```

----End

3.7.2 OS File Permissions

Table 3-20 lists the SUID permissions.

Table 3-20 SUID permissions

File	Functions
/usr/bin/mount	Mounts the file system to a VM.
/usr/bin/umount	Removes the file system from a VM.
/usr/bin/su	Specifies a user or group to run commands.
/usr/bin/chfn	Modifies user finger information.
/usr/bin/chsh	Changes user login shell.
/usr/bin/chage	Modifies the password expiration time and other related information.
/usr/bin/gpasswd	Sets the password of a user group.
/usr/bin/newgrp	Dynamically switches to a new group when users log in to the system.
/usr/bin/crontab	Adds, modifies, and deletes a scheduled task.
/usr/bin/pkexec	Specifies the user identity execution program.

File	Functions
/usr/bin/passwd	Changes a user password.
/usr/bin/sudo	Specifies a user's permission to run a specified command.
/usr/sbin/pam_timestamp_check	Checks whether the default timestamp is valid.
/usr/sbin/unix_chkpwd	Checks password policies.
/usr/sbin/usernetctl	Checks whether the specified user is allowed to operate a network port.
/usr/lib/polkit-1/polkit-agent-helper-1	Serves as an auxiliary command of the polkit agent.
/usr/lib64/dbus-1/dbus-daemon-launch-helper	Starts the dbus daemon process of the system for the standby user.
/usr/sbin/mount.nfs	Provides the NFS client function for installing the NFS.
/usr/bin/fusermount	Mounts and unmounts user space file systems (FUSE).

Table 3-21 lists the SGID permissions.

Table 3-21 SGID permissions

File	Functions
/usr/bin/wall	Sends messages to all user terminals.
/usr/bin/write	Sends messages to another user.
/usr/bin/cgclassify	Moves a specified process to a cgroup.
/usr/bin/cgexec	Executes a specified process in a specified cgroup.
/usr/bin/ssh-agent	Saves private key information during public key authentication.
/usr/sbin/netreport	Sends a SIGIO to the program that invokes this command when the network adapter status changes.
/usr/libexec/utempter/utempter	Assists libutempter to operate the utmp and wtmp files.
/usr/libexec/openssh/ssh-keysign	Generates a digital signature for the host key during host-based authentication.

File	Functions
/usr/bin/screen	Linux built-in software for switching between CLI terminals. Users can establish multiple local or remote CLI sessions at the same time using this software.

4 Service Monitoring

4.1 Overview

Monitoring Purpose

ManageOne is the unified O&M monitoring platform of Huawei Cloud Stack that provides layered monitoring on OSs, middleware, databases, and services, achieving end-to-end fault detection. Huawei Cloud Stack collects metrics and running statuses of monitored hosts, promptly generates alarms on abnormal metrics, and notifies O&M personnel of timely handling to ensure service availability.

Resource Monitoring monitors all resources in the system from underlying physical devices to upper-layer tenant applications to help O&M personnel identify resource risks. If an exception occurs, O&M personnel can view the topology, alarm, and performance data of resources, perform O&M operations such as running logs as well as performing URL tests, and analyze problems step by step to quickly troubleshoot faults.

Monitoring Items

The ManageOne Maintenance Portal provides the following monitoring capabilities for MRS:

- **Management plane monitoring**: The O&M administrator can view the running statuses of MRS service nodes and processes on ManageOne Maintenance Portal.
- **Tenant plane monitoring**: The O&M administrator can view the performance metrics and cluster resource status on ManageOne Maintenance Portal or on the FusionInsight Manager web UI of an MRS cluster, and configure performance metric thresholds for generating alarms.

NOTE

- This document describes only common resource monitoring scenarios related to MRS. For details about the functions and operations related to resource monitoring on ManageOne Maintenance Portal, see "Maintenance Guide" > "O&M Guide" > "Centralized Monitoring" in *Huawei Cloud Stack 8.3.1 Product Documentation* .
- You can view complete monitoring metrics on FusionInsight Manager of the MRS cluster. For details about the metrics, see *MapReduce Service (MRS) 3.3.1-LTS Monitoring Metrics (for Huawei Cloud Stack 8.3.1)*.

4.2 System Resource Monitoring

4.2.1 Viewing Overview Information

Administrators can view MRS information, such as basic information, alarm statistics, performance metrics, component running statuses, automation task statuses, and URL test task statuses on the **Resource Monitoring** page.

Viewing Cloud Service Overview

Step 1 Log in to ManageOne Maintenance Portal and choose **Monitoring** > **Resource Monitoring**. The **Resource Monitoring** page is displayed.

Step 2 Select **Cloud Services** and then **All**, and click the **Cloud Services** tab.

Step 3 Search for **MRS**.

Step 4 In the result list, click **MRS**_*xxx* to go to the MRS monitoring overview page of the corresponding region.

Step 5 View the MRS overview information. For details, see **Table 4-1**.

Table 4-1 Cloud service overview

Card Name	Description
Dashboard	Basic version of MRS in the current region
Automation Task	Statistics on the number of successful and failed automation jobs of the current cloud service
URL Test	Statistics on the number of successful and failed URL test tasks executed by the current cloud service
Alarms	Alarms reported by MRS management-plane and tenant-plane clusters
Component status	Statistics on the alarm status of nodes, microservice instances, and pods where the current component is deployed

Card Name	Description
Key Metrics	Top 5 management VMs are ranked by the percentage of key monitoring metrics in descending order. Key metrics include but are not limited to the following: • CPU Usage • Physical Memory Usage • Disk Usage • Disk I/O Wait • Disk I/O Time Ratio Click **Custom** in the upper right corner of the **Key Metrics** area to select the monitoring time range as needed. Key metrics are preset in the system. You can click **Monitoring Management** to monitor custom metrics.
Monitoring Management	Click **Monitoring Management** to enter the page. After adding a trend chart or top N chart type, return to the **Overview** page to view details about the added monitoring metric. • By selecting objects and metrics, you can customize the trend chart of each metric of the monitored objects. • You can set the following parameters to view the TopN chart of each metric of the monitored objects. – Select the object type. – Metric – Sorting order: ascending or descending. – Aggregation Mode: **avg** (average value), **min** (minimum value), and **max** (maximum value).

----End

Viewing the Overview of VMs on the Management Plane

Step 1 Log in to ManageOne Maintenance Portal and choose **Monitoring** > **Resource Monitoring**.

Step 2 Select **Cloud Resources** and then **Management VMs**. Search for a VM name and click the search result to go to the **Summary** page.

Table 4-2 VMs on the MRS management plane

Service/Component	VM Name	Description
EICommon	EICommon-Region-Master-01 EICommon-Region-Master-02 EICommon-Region-Master-03	Master nodes in the CloudAutoDeploy-CDK cluster in the Region zone
CloudAutoDeploy-CDK cluster	MRS_Region_Node-01 MRS_Region_Node-02 MRS_Region_Node-03	Data nodes in the CloudAutoDeploy-CDK cluster in the Region zone
MRS-DB	MRS_DB-01 MRS_DB-02	MRS database nodes

Step 3 In the navigation pane on the left, choose **Components** to view the NIC and disk information about the VM.

Step 4 In the navigation pane on the left, choose **Monitoring** to view the VM monitoring metrics, such as CPU and memory usage.

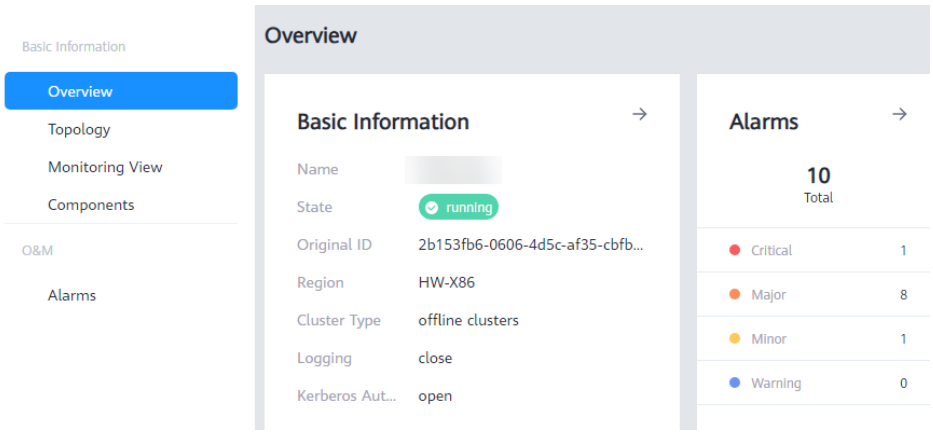
----End

Viewing the Overview of a Tenant Cluster

Step 1 Go to the overview page of MRS by referring to **Viewing Cloud Service Overview**.

Step 2 Click **Tenant Instances** to view the resource summary of all tenants in the current region. Click a cluster name to go to its **Basic Information** page.

Step 3 In the navigation pane on the left, click **Overview** to view **Basic Information** and **Alarms** of the cluster.



----End

4.2.2 Viewing the Topology View

The topology view displays the associations between MRS resources and related resources. O&M personnel can quickly locate faulty resources by viewing the physical and logical topologies of resources.

For details about cloud service topologies, see "Maintenance Guide" > "O&M Guide" > "Centralized Monitoring" in *Huawei Cloud Stack 8.3.1 Product Documentation* .

Accessing the Topology View

Step 1 Log in to ManageOne Maintenance Portal and choose **Monitoring** > **Resource Monitoring**. The **Resource Monitoring** page is displayed.

Step 2 Select **Cloud Services** and then **All**, and click the **Cloud Services** tab.

Step 3 Search for **MRS**.

Step 4 In the result list, click **MRS**_*xxx* to go to the MRS monitoring overview page of the corresponding region.

Step 5 In the navigation pane on the left, choose **Basic Information** > **Topology**.

----End

Viewing the Logical Topology

The logical topology consists of the Global zone and Region zone.

- The Global zone displays the ingress path of service traffic.
- The Region zone displays dependencies between the current service and other cloud services.

A logical topology consists of nodes and links.

- If a node is displayed in green, no alarms are generated on the node. If a node is displayed in red, at least one alarm is generated on the node.

  After you click a node, the basic information, component list, alarm list, URL dialing test, and performance metrics of the node are displayed on the right.

- The dotted box shows the MRS service nodes. Click → in the upper right corner of the dotted box to go to the **Logical Topology** page of MRS service nodes.
- Link: displays the dependency between two resource nodes.
  - The link color is related only to the URL dialing test alarm.
    - If no URL dialing test alarm is reported for a link, the link is displayed in green.
    - If a critical URL dialing test alarm is reported for a link, the link is displayed in red.
    - If a major URL dialing test alarm is reported for a link, the link is displayed in orange.
    - If a minor URL dialing test alarm is reported for a link, the link is displayed in yellow.
    - If a warning URL dialing test alarm is reported for a link, the link is displayed in blue.

    When multiple alarms are generated on a link, the color of the alarm with the highest severity is displayed.
  - The arrow of a link depends on the direction of the URL dialing test case. Typically, an arrow is displayed on the destination node. Because the service logic is multi-directional, a link can be unidirectional or bidirectional.
  - You can click a link to view the list of URL dialing test tasks on the link. The list is obtained by filtering the URL test tasks preset in the current MRS adaptation package based on the source service, source microservice, destination service, and destination microservice. Only the URL test tasks that use one end of the current link as the source and the other end as the destination are displayed.

Viewing the Physical Topology

The physical topology displays logical relationships between the current resource and resources at each level.

- Move your cursor over a resource icon to view the resource name, type, and ID.
- View which resources at each layer are affected when the resource running status is abnormal.
- Double-click a resource icon to go to the **Topology** page. Only details of some resources can be viewed.
- A topology view displays the topology structure from the virtualization layer to the physical layer, that is, from cloud services, services, microservices, microservice instances, deployment nodes, to hosts in sequence.

4.2.3 Viewing Alarms

O&M personnel can view current alarms of resources associated with MRS to obtain resource statuses and handle current alarms in a timely manner.

Procedure

Step 1 Log in to ManageOne Maintenance Portal and choose **Monitoring** > **Resource Monitoring**. The **Resource Monitoring** page is displayed.

Step 2 Select **Cloud Services** and then **All**, and click the **Cloud Services** tab.

Step 3 Search for **MRS**.

Step 4 In the result list, click **MRS**_*xxx* to go to the MRS monitoring overview page of the corresponding region.

Step 5 In the navigation pane, choose **O&M** > **Alarms**.

- Management plane alarms: alarms generated for the deployment nodes of the current cloud service and for the objects (including but not limited to management VMs, host machines, physical servers, and pods) directly associated with the deployment nodes of the current cloud service.
- Tenant plane alarms: alarms reported by clusters provisioned by the current cloud service.
- If there are excessive alarms, you can filter alarms by time range, alarm source, alarm serial number, or alarm ID in the search box.
- In the alarm list, click an alarm name to view its details.

----End

4.2.4 Viewing the Monitoring View

O&M personnel can view the change trend of each compute or storage monitoring metric of MRS.

Viewing Monitoring Metrics of a Microservice

Step 1 Log in to ManageOne Maintenance Portal and choose **Monitoring** > **Resource Monitoring**. The **Resource Monitoring** page is displayed.

Step 2 Select **Cloud Services** and then **All**, and click the **Cloud Services** tab.

Step 3 Search for **MRS**.

Step 4 In the result list, click **MRS**_*xxx* to go to the MRS monitoring overview page of the corresponding region.

Step 5 In the navigation pane, choose **Basic Information** > **Monitoring View**.

Step 6 Select a service, microservice, or microservice instance from the service tree.

Step 7 Select the metric type, metric, time, and dimension to view the change trend of the metric.

NOTE

- Metric types include but are not limited to OS metrics and service metrics.
- Click **Deployment Node** to view details about the nodes deployed under the current resource.

----End

Viewing Monitoring Metrics of a Tenant Cluster

Step 1 Go to the overview page of MRS by referring to **Viewing Cloud Service Overview**.

Step 2 Click **Tenant Instances** to view the resource summary of all tenants in the current region. Click a cluster name to go to its **Basic Information** page.

Step 3 In the navigation pane on the left, choose **Monitoring**.

Step 4 View the monitoring metrics of each resource type in the current cluster.

- **Metric Category**: Select the service whose monitoring metrics you want to view in the current cluster.
- **Period**: You can select different time periods, such as **Recent 1 Hour**, **Recent 3 Hours**, **Recent 12 Hours**, **Recent Day**, or **Recent 3 Days**, to view the monitoring status of the current cluster. You can also customize the time period.

----End

4.2.5 Viewing Components

O&M personnel can view the list of management nodes and microservice instances to which MRS belongs, pay attention to the alarm status and running status of components, and rectify faults in a timely manner. This prevents risks.

Procedure

Step 1 Log in to ManageOne Maintenance Portal and choose **Monitoring** > **Resource Monitoring**. The **Resource Monitoring** page is displayed.

Step 2 Select **Cloud Services** and then **All**, and click the **Cloud Services** tab.

Step 3 Search for **MRS**.

Step 4 In the result list, click **MRS**_*xxx* to go to the MRS monitoring overview page of the corresponding region.

Step 5 In the navigation pane, choose **Basic Information** > **Components**.

Step 6 View details about the components on the management plane.

Table 4-3 MRS component list details

Parameter	Description
Services	Displays information such as the MRS service name and alarm status.
Microservices	Displays information such as the MRS microservice name and alarm status.
Microservice Instances	Displays the name and status of each MRS microservice instance. You can restart microservice instances of the container type.

Parameter	Description
Deployment Nodes	Displays the name and alarm status of the node where MRS is deployed. You can diagnose the deployment node.
Management VMs	Displays the name and status of the MRS management VM.
Host Machines	Displays information about host machines associated with the MRS service.

----End

4.2.6 Viewing Tenant Instances

O&M personnel can view information about all MRS tenant instances, including the running status, to rectify any faults in a timely manner.

Procedure

Step 1 Log in to ManageOne Maintenance Portal and choose **Monitoring** > **Resource Monitoring**. The **Resource Monitoring** page is displayed.

Step 2 Select **Cloud Services** and then **All**, and click the **Cloud Services** tab.

Step 3 Search for **MRS**.

Step 4 In the result list, click **MRS**_*xxx* to go to the MRS monitoring overview page of the corresponding region.

Step 5 In the navigation pane on the left, choose **Basic Information** > **Tenant Instances**.

On the **MapReduce Service** tab page, basic information about all MRS clusters created by the tenant is displayed. You can click a cluster name to view the cluster details and monitoring view.

On the **VMs** tab page, basic information about the ECSs corresponding to the MRS cluster is displayed. You can click a VM name to view its monitoring details.

----End

4.3 Operations and Maintenance

4.3.1 Performing a URL Test

O&M personnel can perform dialing tests on URLs, ports, and IP addresses on the resource management plane associated with the MRS service to ensure resource availability.

This section describes how to create and execute a URL test task for MRS. For details about more functions and operations, see "Maintenance Guide" > "O&M

Guide" > "Centralized Monitoring" in *Huawei Cloud Stack 8.3.1 Product Documentation* .

Viewing and Starting a URL Test

Step 1 Log in to ManageOne Maintenance Portal and choose **Monitoring** > **Resource Monitoring**. The **Resource Monitoring** page is displayed.

Step 2 Select **Cloud Services** and then **All**, and click the **Cloud Services** tab.

Step 3 Search for **MRS**.

Step 4 In the result list, click **MRS**_*xxx* to go to the MRS monitoring overview page of the corresponding region.

Step 5 In the navigation pane on the left, choose **O&M** > **URL Test**.

On the **URL Test (Management Plane)** tab page that is displayed, test cases include preset cases in the current cloud service adaptation package and custom test cases that belong to the current cloud service.

Step 6 Click **Start** to start the URL test task.

Table 4-4 lists the parameters for creating a test task.

Table 4-4 Kye parameters in the URL test task list

Parameter	Description
Task Name	Name of a URL test task Click the name of a test task to view its details.
Test Scenario	Scenario of a dialing test task, including fault locating and fault detection.
Agreement	URL dialing test type
URL	URL of the resource to be tested
Task Status	Status of a URL test task, which can be **Running**, **Failed**, **Stopped**, or **Starting**.
Result Status	Execution result of the current URL test task, which can be **Successful**, **Timeout**, or **Failed**
Interval	Interval for performing a URL test task
Response Time	Interval for performing a URL test task
Resource Name	Name of a cloud service that initiates a URL test task
Test Point	Region to which the cloud service that initiates a URL test task belongs
Region	Name of a region where the URL test point is located

Parameter	Description
Alarm Reported	If alarm parameters are set during URL test task creation, this parameter is set to **Yes**. Otherwise, this parameter is set to **No**.
Preset	Whether a URL test task is preset in the system. Preset tasks cannot be modified or deleted.
Operation	You can perform the following operations on a test task: • **Test**: Click **Test**. If "Test Success" is displayed in the upper right corner, the task passed the test. • **Start**/**Stop**: Start or stop a URL test task. • **Modify**: Modify parameters of a URL test task. • **Delete**: Delete a URL test task.

Click the name of a test task to view its details.

----End

Creating a URL Test Task

Step 1 On the **URL Test (Management Plane)** tab page, click **Create** to create a URL test task. Set the parameters according to **Table 4-5**.

Table 4-5 Parameters for creating a URL test task

Parameter		Description	Example Value
Items	Name	Name of a URL test task	Payment
	Type	**HTTP**, **HTTPS**, **ICMP**, and **TCP** are supported.	HTTPS
	Address	URL to be tested Set this parameter as prompted.	https://www.example.com:8080/healthcheck
	Authentication Mode	URL authentication mode, which can be **IAM** or **No Auth**. Tests whose **Type** is **ICMP** or **TCP** do not involve authentication.	IAM

Parameter		Description	Example Value
	Request Mode	Standard HTTP or HTTPS request methods include GET, POST, and HEAD. The POST method can be used to submit content. Tests whose **Type** is **ICMP** or **TCP** do not involve request modes.	GET
	Advanced Setting	For details, see **Table 4-6**. Tests whose **Type** is **ICMP** or **TCP** do not involve advanced configurations.	-
	Frequency	Interval for performing a URL test task. This parameter has a fixed value of **10min**.	10min
Test Object	Resource	Select the cloud service resource to which the URL belongs.	MRS_*XXX*
	Test Point	The test point is an ECS. You can select an existing ECS or create one. You are advised to create an ECS dedicated for a test as the test point.	-
Alarm	Enable Alarms	Select **Yes** or **No**. This parameter is not available for tests whose **Type** is **ICMP** or **TCP**.	Yes
	Number of Consecutive	An alarm is reported only when the number of consecutive times that the alarm is generated reaches a specified value.	2
	Severity	Alarm severity, which can be **Critical**, **Major**, **Minor**, or **Warning**.	Major

Table 4-6 HTTP(S) advanced parameters

Parameter	Description	Example Value
Request Content	The content to be submitted by the application to be tested when the request mode is POST.	{"username":"wang"}

Parameter	Description	Example Value
HTTP Request Headers	HTTP header information required by the application to be tested. **NOTE** If the submitted content is in the JSON format, the value of **HTTP Request Header** must be **Content-Type: application/json**.	key=Content-Type, value= application/json
Request Cookies	Cookies required by the application to be tested	key=Hm_lvt_e7a90fbb*xxxx*0aec64d1170a5ca608f, value=1634402786
Match Response Content	When the request mode is GET or POST, set the response body matching mode. • **Yes**: If the response body contains the specified matched string, the test succeeds. • **No**: If the response body does not contain the specified matched string, the test succeeds.	Yes
Response Content to Match	Expected HTTP response content when the request mode is GET or POST.	-

Step 2 Click **Create Now**.

- To test the task, click **Test**.
- To cancel the creation, click **Cancel**.

----End

4.3.2 Run Logs

4.3.2.1 Collecting Logs Using a Log Template

Template logs are templates for collecting MRS run logs. Some log templates are preset in the system and can be downloaded by administrators as needed. If the preset log templates cannot meet service requirements, administrators can create and download new log templates.

Collecting Logs Using a Log Template

Step 1 Log in to ManageOne Maintenance Portal and choose **Monitoring** > **Resource Monitoring**. The **Resource Monitoring** page is displayed.

Step 2 Select **Cloud Services** and then **All**, and click the **Cloud Services** tab.

Step 3 Search for **MRS**.

Step 4 In the result list, click **MRS**_*xxx* to go to the MRS monitoring overview page of the corresponding region.

Step 5 In the navigation pane, choose **O&M** > **Run Logs**.

Step 6 Choose **Management Plane** > **Template Logs**. In the displayed log template list, you can view all preset log templates.

Step 7 Click **Create Template** to create a log template.

1. Set **Name**.
2. Set **Description**.
3. In the **Add** area, select a cloud service, service or microservice, and path for storing logs.
   - To add multiple log files, click **Add**. A maximum of 15 log files can be added to a template.
   - To delete an added log path, click **Delete** in the **Operation** column.

Step 8 Click **OK**.

Click a template name. On the **Template Details** page that is displayed, view **Cloud Service**, **Service/Microservice**, and **Path** in the log template.

Step 9 Click **Download** in the **Operation** column. On the **Create Download Task** page that is displayed, create a download task.

1. Set **Task Name**.
2. Set the period for the log file. Click **Custom** next to **Period** to customize the period for the log to be downloaded.

Step 10 Click **OK**. The **Download Tasks** tab page is displayed, where you can view the task progress. To download logs to your local PC, click **Download** in the **Operation** column.

----End

Downloading Logs to the Local PC

After a log template download task is created, a task list is automatically generated on the **Management Plane** > **Download Tasks** page. You can download logs to the local PC.

Step 1 Click the **Download Tasks** tab to view all created log template download tasks.

Table 4-7 Parameters of created log template download tasks

Parameter	Description
Name	Name of a log download task
Username	User who downloads logs
Log Package Size	Size of downloaded logs **NOTE** • A single log package to be downloaded cannot exceed 2 GB. • A single log file to be downloaded cannot exceed 100 MB.
Time Range	Time required for downloading logs

Parameter	Description
Created	Time when a log download task is created
Status	Execution status of the current task. If **Successful** is displayed, the task is successfully executed. **NOTE** All log files generated within the selected time range will be downloaded. All the files that meet the requirements in the last-level directory are downloaded. If **Status** in the task list is **Failed. There are failed subtasks.** and the failure cause of the subtask is **Task file download timed out.**, perform the following operations: 1. Click **Retry** to view the task status in the task list. – **Successful**: The task is complete. – If the download fails, create a download task 40 minutes later, and check the task status. For details, see "Collecting Logs Using a Log Template." 2. If the download still fails, contact technical support for assistance.
Progress	Execution progress of the current task. If **Progress** reaches 100%, the logs are collected.

Step 2 In the upper right corner of the list, set the time range and filter criteria to filter required tasks based on service requirements.

NOTE

You can view the tasks only in last two days. Tasks that have been created for more than two days are automatically deleted.

Step 3 Click **Download** in the **Operation** column to download the log file to the local PC for analysis.

----End

4.3.2.2 Obtaining Logs of a Specified Node

When locating cloud service faults, O&M personnel need to view management plane run logs of cloud services and download these logs to the local PC for analysis.

Procedure

Step 1 Log in to ManageOne Maintenance Portal and choose **Monitoring** > **Resource Monitoring**. The **Resource Monitoring** page is displayed.

Step 2 Select **Cloud Services** and then **All**, and click the **Cloud Services** tab.

Step 3 Search for **MRS**.

Step 4 In the result list, click **MRS**_*xxx* to go to the MRS monitoring overview page of the corresponding region.

Step 5 In the navigation pane, choose **O&M** > **Run Logs**.

Step 6 Choose **Node Logs**.

Step 7 In the left pane, you can select **By Microservice** or **By Node** to select the nodes whose log files need to be downloaded. The log files of the selected nodes are dynamically displayed in the download list on the right.

NOTE

- A maximum of five nodes can be selected at a time.
- A maximum of 100 logs can be viewed on a single node.
- Nodes cannot be selected across services or microservices.

Table 4-8 Parameters in the node log list

Parameter	Description
Log Name	Name of a log file
File Size	Size of a log file
Last Modified	Time when a log file was modified for the last time
Log Path	Path for storing log files
Management VM	Management VM to which a log file belongs
Service/Microservice	Service or microservice to which the log file belongs

Step 8 In the upper right corner of the page, set the time range and filter criteria to select required log files.

NOTE

You can select up to seven days as the time range.

Step 9 Download node logs on the management plane.

- To download log files in batches, select the log files to be downloaded and click **Download** above the log list. You can select log files on different pages.
- To download a single log, click **Download** in the **Operation** column.

  NOTE

  - A maximum of 20 logs can be downloaded at a time.
  - A single log file to be downloaded cannot exceed 100 MB, and the total size of all log files to be downloaded cannot exceed 2 GB.

----End

4.3.3 Manually Enabling MRS Cluster Performance Metric Reporting

Scenario

Related metrics can be reported to ManageOne Maintenance Portal only after an agency is configured and metric sharing is enabled during MRS cluster creation. If

the function is not enabled during MRS cluster creation, you can manually enable it by following the instructions provided in this section.

Procedure

Step 1 Log in to the MRS console. On the **Active Clusters** page, click a cluster name.

Step 2 On the **Dashborad** tab page of the cluster details page, click **Manage Agency** on the right of **Agency** and select the agency to be bound to the MRS cluster. Bind an agency with at least the **ces admin** permission to the MRS cluster.

Step 3 Log in to the Master1 node as user **root**.

Step 4 Run the following commands to modify and view the parameters in the installation script:

vim /opt/Bigdata_func/cloudinit/cloudinit_params

Change **cesMonitor=false** to **cesMonitor=true** and save the configuration.

Check whether the domain names corresponding to the values of **obs_domain_name** and **ces_endpoint** in the **cloudinit_params** script can be pinged. If the domain names cannot be pinged, contact O&M engineers.

Step 5 Run the following command to run the installation script:

sh /opt/Bigdata_func/cloudinit/install_metric_agent.sh

After the script is executed, search for **send** in the **/var/log/metric-agent/log/metric-agent.log** file. If the interface is 200, the reporting is successful. You can view the reported performance indicator data in the ManageOne Maintenance Portal file.

----End

Additional Information

- After the installation, the following information is displayed:
  ```
  kinit: Client '1@HADOOP.COM' not found in Kerberos database while getting initial credeals
  ```
  This issue does not affect functions and can be ignored.
- After the installation, the following information is displayed:
  package metric-agent-1.1.0-1.x86_64 is already installed or **package metric-agent-1.1.0-1.aarch64 is already installed**

  The RPM package has been installed on the node. Run the **sh /opt/Bigdata_func/cloudinit./install_metric_agent.sh** command after uninstalling the package.
- To uninstall **metric-agent**, perform the following steps:
  a. Run the **rpm -qa metric*** command on the Master1 node to check whether the **metric-agent** has been installed.
  b. Run the following command to uninstall the RMP package: (You can run the **arch** command to view the CPU architecture of the current node.)

     rpm -e metric-agent-1.1.0-1.x86_64 or **rpm -e metric-agent-1.1.0-1.aarch64**

     Ignore the following information during the uninstallation:
     ```
     warning: file /tmp/metric-agent.tar.gz: remove failed: No such file or directory
     ```


c. Delete related files and directories.

rm -rf /usr/local/metric-agent

rm -rf /var/log/metric-agent

Stop the process.

ps -ef | grep MetricAgent

kill -9 *Process ID*

4.4 Monitoring Metrics

You can view complete monitoring metrics on FusionInsight Manager of an MRS cluster. For details about the metrics, see *MapReduce Service (MRS) 3.3.1-LTS Monitoring Metrics (for Huawei Cloud Stack 8.3.1)*.

Viewing Monitoring Metrics of a Cluster on the MRS Console

Step 1 Log in to ManageOne as the VDC administrator or VDC operator using a browser.

- URL in non-B2B scenarios: **https://***Domain name of ManageOne Operation Portal*, for example, **https://console.demo.com**
- URL in B2B scenarios: **https://***Domain name of ManageOne Tenant Portal*, for example, **https://tenant.demo.com**
- URL of the unified portal: **https://***ManageOne Unified Portal URL*, for example, **https://console.demo.com/moserviceaccesswebsite/unifyportal#/home**. On the home page, choose **Self-service Cloud Service Center**.

Step 2 Click ☰ in the upper left corner of the page and select a region. Then choose **EI Enterprise Intelligence** > **MapReduce Service**.

Step 3 Choose **Clusters** > **Active Clusters** and click a cluster name to go to its details page.

Step 4 In the O&M management area of the **Dashboard** tab page, click **Synchronize** next to **IAM User Sync** to synchronize IAM users.

Step 5 After the synchronization is complete, you can view the cluster monitoring metric report on the **Monitor** page.

Step 6 Specify a period to view monitoring data.

Step 7 Customize a monitoring report.

1. Click **Customize** and select monitoring metrics to be displayed.
2. Click **OK** to save the selected monitoring metrics for display.

Step 8 Click the **Components** tab and then a service name to view the detailed metrics of the current service.

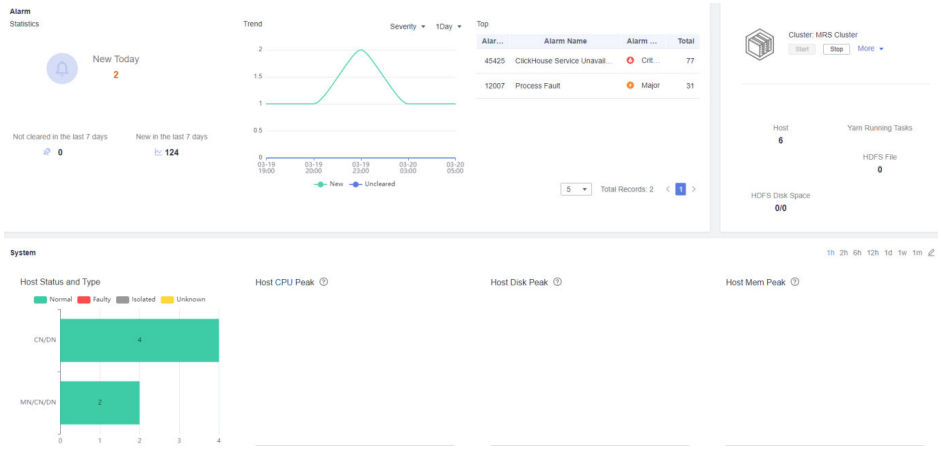
----End

Viewing Monitoring Metrics of a Cluster on FusionInsight Manager

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

Step 2 Click **Homepage** to view the main monitoring metric report of the cluster.

Step 3 In the upper right corner, specify a period to view monitoring data.



Step 4 Customize a monitoring report.

1. In the upper right corner, click , select **Customize**, and select monitoring metrics to be displayed.
2. Click **OK** to save the selected monitoring metrics for display.

Step 5 In the navigation pane on the left, choose **Cluster** > **Services** and click a component name to view the detailed metrics of the current service.

----End

5 Critical Operations

5.1 High-Risk Operations on the Management Plane

During system O&M, you must strictly follow the operation guide when performing the risky operations listed in the following table. Otherwise, potential risks may arise, affecting proper system running.

High-risk operations are classified into **Table 5-1** and **Table 5-2**.

Table 5-1 Common basic operations

Operation Object	Operation	Risk	Severity	Recommendation
Operating System (OS)	Operations on files and directories	Deleting the entire directory and file using **rm**	High	1. Bastion host management permission is used so that non-O&M personnel cannot log in to the system. 2. The bastion host displays risky commands. Exercise caution when performing this operation. 3. The following operations are forbidden. If the following operations are required, they must be approved and performed by the professional O&M personnel. One person performs the operations and the other supervises the operations. a. Deleting or moving a file or directory b. Start or stop the system or service processes without permission, including containers and processes in the containers. c. Modifying OS parameters or other configurations 4. In the modification scenario, the change guide and test report must be provided. The professional O&M personnel are responsible for the operation, and the operation must be supervised.
		Moving files and directories using **mv**.	High	
		Changing the owner and permission of the file or directory using the **chmod** and **chown** commands	High	
		Directly compressing files using compression commands such as **gzip**, **bzip2**, or **compress**	High	
		Using * instead of specified files when running the **grep**, **tar cvzf**, **find**, or **ls** commands.	Medium	
	Process operations	Stopping and starting the process	High	
		Modifying process startup parameters, for example, add environment variables	High	
	OS parameters	Shutting down the operating system using **shutdown**	High	
		Restarting the operating system using **reboot**	High	

Operation Object	Operation	Risk	Severity	Recommendation
		Adding, deleting, and modifying IP addresses, routes, or host names.	High	
		Modifying firewall rules	High	
		Modifying the **etc/host** file	High	
		Modifying DNS parameters	High	
		Modifying NTP parameters	High	
		Attaching a hard disk	High	
		Restarting or stopping the service network adapter	High	

Operation Object	Operation	Risk	Severity	Recommendation
		Modifying the kernel parameters of the operating system Example: 1. Changing the maximum number of processes used by each user, modifying I/O parameters, or changing the maximum memory space occupied by the file system 2. Modifying the maximum number of file handles, maximum shared memory, or core file size 3. Core file size	High	
		Formatting the partition of the running operating system	High	
		Modifying network services such as SSH or SFTP	High	
	Account operations	Changing the password of the **root** user or the default service user	High	
		Deleting or adding a user	High	

Operation Object	Operation	Risk	Severity	Recommendation
	Docker&K8S	Deleting all pods of a service at the same time (**kubectl delete pods**).	High	

Table 5-2 Operations on the MRS management plane

Operation Object	Operation	Risk	Severity	Recommendation
Console	Global Console	Modifying Tomcat configuration files, for example, **server.xml**	Low	1. A formal test report and a change guide are required for each change. Risky operations must be clearly described in the change guide, and rollback measures must be provided. 2. The change must be performed by professional O&M personnel. One person is responsible for the change, and the other supervises the operation. 3. If API changes and parameter changes are involved, risk assessment is required before the change review.
		Changing IP address	High	
		Modifying the services associated with a region	High	
		Adjust the console framework, third-party dependency components (such as JSON parsing and encryption/ decryption components), and console functions during the console upgrade or patch installation.	High	
	Region Console	Modifying Tomcat configuration files, for example, **server.xml**	High	

Operation Object	Operation	Risk	Severity	Recommendation
		Changing IP address	High	
		Adjust the console framework, third-party dependency components (such as JSON parsing and encryption/ decryption components), and console functions during the console upgrade or patch installation.	High	
Service function/ component changes	Version upgrading	Modifying peripheral interfaces of a version	High	
		Modifying the API of a version	High	
		Modifying the method of referencing third-party APIs for a service	High	
	Taking components offline	Replacing third-party components of a service	High	
		Taking service components offline	High	

5.2 High-Risk Operations on the Tenant Plane

Forbidden Operations

Table 5-3 lists forbidden operations during the routine cluster operation and maintenance process.

Table 5-3 Forbidden operations

Item	Risk
Delete ZooKeeper data directories.	ClickHouse, HDFS, Yarn, HBase, and Hive depend on ZooKeeper, which stores metadata. This operation has adverse impact on normal operating of related components.
Frequently switch over the active and standby JDBCServer nodes.	This operation may interrupt services.
Delete Phoenix system tables and data (SYSTEM.CATALOG, SYSTEM.STATS, SYSTEM.SEQUENCE, and SYSTEM.FUNCTION).	This operation will cause service operation failures.
Manually modify data in the Hive metabase (hivemeta database).	This operation may cause Hive data parse errors. As a result, Hive cannot provide services.
Change permission on the Hive private file directory **hdfs:///tmp/hive-scratch**.	This operation may cause unavailable Hive services.
Modify **broker.id** in the Kafka configuration file.	This operation may cause invalid node data.
Reduce the capacity of a Redis cluster that contains large values.	This operation will cause a failure to reduce the capacity of the Redis cluster.
Modify the host names of nodes.	Instances and upper-layer components on the host cannot provide services properly. The fault cannot be rectified.
Delete or update data in the database as a HetuEngine user.	The HetuEngine compute instance cannot be used.

The following tables list the high-risk operations during the operation and maintenance of each component.

Manager High-Risk Operations

Table 5-4 Manager high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Change the OMS password.	This operation will restart all processes of OMSServer, which has adverse impact on cluster maintenance and management.	▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check whether there are uncleared alarms and whether the cluster management and maintenance are normal.
Import the certificate.	This operation will restart OMS processes and the entire cluster, which has adverse impact on cluster maintenance and management and services.	▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Perform an upgrade.	This operation will restart Manager and the entire cluster, affecting management, maintenance, and services of the cluster. Strictly manage the user who is eligible to assign the cluster management permission to prevent security risks.	▲ ▲ ▲	Ensure that there is no other maintenance and management operations when the operation is performed.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.

Operation	Risk	Severity	Workaround	Check Item
Restore the OMS.	This operation will restart Manager and the entire cluster, affecting management, maintenance, and services of the cluster.	▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Change an IP address.	This operation will restart Manager and the entire cluster, affecting management, maintenance, and services of the cluster.	▲ ▲ ▲	Ensure that there is no other maintenance and management operations when the operation is performed and that the new IP address is correct.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Change log levels.	If the log level is changed to **DEBUG**, Manager responds slowly.	▲ ▲	Before the modification, confirm the necessity of the operation and change it back to the default log level in time.	None

Operation	Risk	Severity	Workaround	Check Item
Replace a control node.	This operation will interrupt services deployed on the node. If the node is a management node, the operation will restart all OMS processes, affecting the cluster management and maintenance.	▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Replace a management node.	This operation will interrupt services deployed on the node. As a result, OMS processes will be restarted, affecting the cluster management and maintenance.	▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Restart the upper-layer service at the same time during the restart of a lower-layer service.	This operation will interrupt the upper-layer service, affecting the management, maintenance, and services of the cluster.	▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.

Operation	Risk	Severity	Workaround	Check Item
Modify the OLDAP port.	This operation will restart the LdapServer and Kerberos services and all associated services, affecting service running.	▲ ▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	None
Delete the **supergroup** group.	Deleting the **supergroup** group decreases user rights, affecting service access.	▲ ▲ ▲ ▲ ▲	Before the change, confirm the rights to be added. Ensure that the required rights have been added before deleting the **supergroup** rights to which the user is bound, ensuring service continuity.	None

Operation	Risk	Severity	Workaround	Check Item
Reinstall a host.	This operation will reinstall the software on the specified host and may cause data loss due to the cleanup of the data directory.	▲ ▲ ▲	Before performing this operation, ensure that the reinstallation is necessary and exercise caution when selecting the data cleanup option. **NOTE** This function applies only to MRS physical machine clusters.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Reinstall an instance.	This operation will reinstall the instance on the specified host and may cause data loss due to the cleanup of the data directory.	▲ ▲ ▲	Before performing this operation, ensure that the reinstallation is necessary and exercise caution when selecting the data cleanup option. **NOTE** This function applies only to MRS physical machine clusters.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.

Operation	Risk	Severity	Workaround	Check Item
Restart a service.	Services will be interrupted during the restart. If you select and restart the upper-layer service, the upper-layer services that depend on the service will be interrupted.	▲ ▲ ▲	Confirm the necessity of restarting the system before the operation.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Change the default SSH port No.	After the default port (22) is changed, functions such as cluster creation, service/instance adding, host adding, and host reinstallation cannot be used, and results of cluster health check items for node mutual trust, **omm**/**ommdba** user password expiration, and others are incorrect.	▲ ▲ ▲	Before performing this operation, restore the SSH port to the default value.	None

Operation	Risk	Severity	Workaround	Check Item
Power off and power on a system.	If a system is powered off and powered on in a non-standard mode, cluster startup will become faulty. For example, LDAP data fails to be synchronized, or controller startup fails.	▲ ▲ ▲	For details about how to power off and power on a system, see "System Power-On and Power-Off" in *MapReduce Service (MRS) 3.3.1-LTS User Guide (for Huawei Cloud Stack 8.3.1) in the MapReduce Service (MRS) 3.3.1-LTS Usage Guide (for Huawei Cloud Stack 8.3.1)*. **NOTE** This function applies only to MRS physical machine clusters.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.

CDL High-risk Operations

Table 5-5 CDL high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Start or stop basic components independently.	This operation has adverse impact on the basic functions of some services. As a result, service failures occur.	▲ ▲ ▲	Do not start or stop basic components such as Kafka, DBService, ZooKeeper, Kerberos, and LDAP separately. To start or stop basic components, select associated services.	Check whether the service status is normal.
Restart or stop services.	This operation may interrupt services.	▲ ▲	Restart or stop services when necessary.	Check whether the service is running properly.

ClickHouse High-Risk Operations

Table 5-6 ClickHouse high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Delete data directories.	This operation may cause service information loss.	▲ ▲ ▲	Do not delete data directories manually.	Check whether data directories are normal.

Operation	Risk	Severity	Workaround	Check Item
Remove ClickHouseServer instances.	The ClickHouseServer instance nodes in the same shard must be removed in at the same time. Otherwise, the topology information of the logical cluster is incorrect. Before performing this operation, check the database and data table information of each node in the logical cluster and perform scale-in pre-analysis to ensure that data is successfully migrated during the scale-in process to prevent data loss	▲ ▲ ▲ ▲ ▲	Before scale-in, collect information in advance to learn the status of the ClickHouse logical cluster and instance nodes.	Check the ClickHouse logical cluster topology information, database and data table information in each ClickHouseServer instance, and data volume.
Add ClickHouseServer instances.	When performing this operation, you must check whether a database or data table with the same name as that on the old node needs to be created on the new node. Otherwise, subsequent data migration, data balancing, scale-in, and decommissioning will fail.	▲ ▲ ▲ ▲ ▲	Before scale-out, confirm the function and purpose of new ClickHouseServer instances and determine whether to create related databases and data tables.	Check the ClickHouse logical cluster topology information, database and data table information in each ClickHouseServer instance, and data volume.

Operation	Risk	Severity	Workaround	Check Item
Decommission ClickHouseServer instances.	The ClickHouseServer instance nodes in the same shard must be decommissioned in at the same time. Otherwise, the topology information of the logical cluster is incorrect. Before performing this operation, check the database and data table information of each node in the logical cluster and perform decommissioning pre-analysis to ensure that data is successfully migrated during the decommissioning process to prevent data loss	▲ ▲ ▲ ▲ ▲	Before decommissioning, collect information in advance to learn the status of the ClickHouse logical cluster and instance nodes.	Check the ClickHouse logical cluster topology information, database and data table information in each ClickHouseServer instance, and data volume.
Recommission ClickHouseServer instances.	When performing this operation, you must select all nodes in the original shard. Otherwise, the topology information of the logical cluster is incorrect.	▲ ▲ ▲ ▲ ▲	Before recommissioning, you need to confirm the home information about the shards of the node to be recommissioned.	Check the ClickHouse logical cluster topology information.
Modify data directory content (file and folder creation).	This operation may cause the ClickHouse instance of the node faults.	▲ ▲ ▲	Do not create or modify files or folders in the data directories manually.	Check whether data directories are normal.

Operation	Risk	Severity	Workaround	Check Item
Start or stop basic components independently.	This operation has adverse impact on the basic functions of some services. As a result, service failures occur.	▲ ▲ ▲	Do not start or stop ZooKeeper, Kerberos, and LDAP basic components independently. Select related services when performing this operation.	Check whether the service status is normal.
Restart or stop services.	This operation may interrupt services.	▲ ▲	Restart or stop services when necessary.	Check whether the service is running properly.

Containers High-Risk Operations

Table 5-7 Containers high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Delete a BLU.	Services may be interrupted.	▲ ▲ ▲ ▲	Ensure that this operation is necessary.	None
Delete a BLU instance.	This operation increases the service data processing pressure.	▲	Ensure that this operation is necessary.	None
Delete the WebContainer_*N* instance.	Services may be interrupted.	▲ ▲ ▲ ▲	Ensure that this operation is necessary. **NOTE** This function applies only to MRS physical machine clusters.	None

Operation	Risk	Severity	Workaround	Check Item
Update a configuration set.	The service processing logic changes.	▲ ▲	Ensure that this operation is necessary. Back up the old configuration set for rollback.	Check whether services are restored.
Restart a service.	Services are interrupted.	▲ ▲	Ensure that this operation is necessary.	Check whether services are restored.
Stop or restart a container.	Services are interrupted.	▲ ▲	Ensure that this operation is necessary.	Check whether services are restored.

DBService High-Risk Operations

Table 5-8 DBService high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Change the DBService password.	The services need to be restarted for the password change to take effect. The services are unavailable during the restart.	▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check whether there are uncleared alarms and whether the cluster management and maintenance are normal.

Operation	Risk	Severity	Workaround	Check Item
Restore DBService data.	After the data is restored, the data generated after the data backup and before the data restoration is lost. After the data is restored, the configuration of the components that depend on DBService may expire and these components need to be restarted.	▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check whether there are uncleared alarms and whether the cluster management and maintenance are normal.
Delete DBService instances.	Data may be lost and cannot be restored, service running may be faulty, or other instance configurations may expire.	▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time. **NOTE** This function applies only to MRS physical machine clusters.	Check whether there are uncleared alarms and whether the cluster management and maintenance are normal.

Operation	Risk	Severity	Workaround	Check Item
Perform active/standby DBService switchover.	During the DBServer switchover, DBService is unavailable.	▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	None
Change the DBService floating IP address.	The DBService needs to be restarted for the change to take effect. The DBService is unavailable during the restart. If the floating IP address has been used, the configuration will fail, and the DBService will fail to be started.	▲ ▲ ▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.

Flink High-Risk Operations

Table 5-9 Flink high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Change log levels.	If the log level is modified to DEBUG, the task running performance is affected.	▲ ▲	Before the modification, confirm the necessity of the operation and change it back to the default log level in time.	None

Operation	Risk	Severity	Workaround	Check Item
Modify file permissions.	Tasks may fail.	▲ ▲ ▲	Confirm the necessity of the operation before the modification.	Check whether related service operations are normal.

Flume High-Risk Operations

Table 5-10 Flume high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify the Flume instance start parameter **GC_OPTS**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Change the default value of **dfs.replication** from **3** to **1**.	This operation will have the following impacts: 1. The storage reliability deteriorates. If the disk becomes faulty, data will be lost. 2. NameNode fails to be restarted, and the HDFS service is unavailable.	▲ ▲ ▲ ▲	When modifying related configuration items, check the parameter description carefully. Ensure that there are more than two replicas for data storage.	Check whether the default replica number is not 1 and whether the HDFS service is normal.

FTP-server High-Risk Operations

Table 5-11 FTP-server high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify the FTP-Server instance start parameter **GC_OPTS**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Change the value of FTP-Server instance parameter **ftp-server-ip**.	Services cannot start properly.	▲ ▲	Change the IP address based on the actual environment.	Check whether services can be started properly.

HBase High-Risk Operations

Table 5-12 HBase high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify encryption configuration. • hbase.regionserver.wal.encryption • hbase.crypto.keyprovider.parameters.uri • hbase.crypto.keyprovider.parameters.encryptedtext	Services cannot start properly.	▲ ▲ ▲ ▲	Strictly follow the prompt information when modifying related configuration items, which are associated. Ensure that new values are valid.	Check whether services can be started properly.

Operation	Risk	Severity	Workaround	Check Item
Change the value of **hbase.regionserver.wal.encryption** to **false** or switch encryption algorithm from AES to SMS4.	This operation may cause start failures and data loss.	▲ ▲ ▲ ▲	When HFile and WAL are encrypted using an encryption algorithm and a table is created, do not close or switch the encryption algorithm randomly. If an encryption table (ENCRYPTION=>AES/SMS4) is not created, you can only switch the encryption algorithm.	None
Modify HBase instance start parameter **GC_OPTS** and **HBASE_HEAPSIZE**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid. GC_OPTS does not conflict with HBASE_HEAPSIZE.	Check whether services can be started properly.
Use **OfflineMetaRepair** tool	Services cannot start properly.	▲ ▲ ▲ ▲	This tool can be used only when HBase is offline and cannot be used in data migration scenarios.	Check whether HBase services can be started properly.

HDFS High-Risk Operations

Table 5-13 HDFS high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Change HDFS NameNode data storage directory **dfs.namenode.name.dir** and DataNode data configuration directory **dfs.datanode.data.dir**.	Services cannot start properly.	▲ ▲ ▲ ▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Use the **-delete** parameter when you run the **hadoop distcp** command.	During DistCP copying, files that do not exist in the source cluster but exist in the destination cluster are deleted from the destination cluster.	▲ ▲	When using DistCP, determine whether to retain the redundant files in the destination cluster. Exercise caution when using the **-delete** parameter.	After DistCP copying is complete, check whether the data in the destination cluster is retained or deleted according to the parameter settings.

Operation	Risk	Severity	Workaround	Check Item
Modify the HDFS instance start parameter **GC_OPTS**, **HADOOP_HEAPSIZE**, and **GC_PROFILE**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid. GC_OPTS does not conflict with HADOOP_HEAPSIZE.	Check whether services can be started properly.
Change the default value of **dfs.replication** from **3** to **1**.	This operation will have the following impacts: 1. The storage reliability deteriorates. If the disk becomes faulty, data will be lost. 2. NameNode fails to be restarted, and the HDFS service is unavailable.	▲ ▲ ▲ ▲	When modifying related configuration items, check the parameter description carefully. Ensure that there are more than two replicas for data storage.	Check whether the default replica number is not 1 and whether the HDFS service is normal.
Change the remote procedure call (RPC) channel encryption mode (**hadoop.rpc.protection**) of each module in Hadoop.	This operation causes service faults and service exceptions.	▲ ▲ ▲ ▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether HDFS and other services that depend on HDFS can properly start and provide services.

HetuEngine High-Risk Operations

Table 5-14 HetuEngine high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Delete all HSBroker instances.	All HetuEngine compute instances are deleted, information about interconnected data sources is lost, and services are unavailable.	▲▲▲▲	Before deleting HSBroker instances, ensure that at least one HSBroker instance exists.	Check whether the service is running properly and whether interrupted operations are restored.
Modify the configuration items in the configuration files of the HetuEngine compute instance. The configuration files include the **coordinator.config.properties** and **worker.config.properties** files.	The HetuEngine compute instance fails to be started.	▲▲▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether the service is running properly.
Modify the configuration items in the JVM files of the HetuEngine compute instance. The JVM files include the **coordinator.jvm.config** and **worker.jvm.config** files.	The HetuEngine compute instance fails to be started.	▲▲▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether the service is running properly.

Operation	Risk	Severity	Workaround	Check Item
Delete the tenant who has started a HetuEngine compute instance.	The HetuEngine compute instance cannot be used.	▲▲▲	Before deleting a tenant, ensure that the HetuEngine compute instance corresponding to the tenant is no longer used and delete the HetuEngine compute instance on HSConsole.	Check whether the service is running properly.
Restart Yarn when the HetuEngine compute instance is started and the auto scaling function is enabled.	The auto scaling function of the HetuEngine compute instance is unavailable.	▲▲▲	Before restarting Yarn, disable the auto scaling function of HetuEngine compute instance on HSConsole.	Check whether the service is running properly.

Hive High-Risk Operations

Table 5-15 Hive high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify the Hive instance start parameter **GC_OPTS**.	This operation may cause Hive instance start failures.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Delete all MetaStore instances.	This operation may cause Hive metadata loss. As a result, Hive cannot provide services.	▲ ▲ ▲	Do not perform this operation unless ensure that Hive table information can be discarded.	Check whether services can be started properly.
Delete or modify files corresponding to Hive tables over HDFS interfaces or HBase interfaces.	This operation may cause Hive service data loss or tampering.	▲ ▲	Do not perform this operation unless ensure that the data can be discarded or that the operation meets service requirements.	Check whether Hive data is complete.

Operation	Risk	Severity	Workaround	Check Item
Delete or modify files corresponding to Hive tables or directory access permission over HDFS interfaces or HBase interfaces.	This operation may cause related service scenarios to be unavailable.	▲ ▲ ▲	Do not perform this operation.	Check whether related service operations are normal.
Delete or modify **hdfs:///apps/templeton/hive-3.1.0.tar.gz** over HDFS interfaces.	WebHCat fails to perform services due to this operation.	▲ ▲	Do not perform this operation.	Check whether related service operations are normal.
Export table data to overwrite the data at the local. For example, export the data of **t1** to **/opt/dir**. **insert overwrite local directory '/opt/dir' select * from t1;**	This operation will delete target directories. Incorrect setting may cause software or OS startup failures.	▲ ▲ ▲ ▲ ▲	Ensure that the path where the data is written does not contain any files or do not use the key word overwrite in the command.	Check whether files in the target path are lost.

Operation	Risk	Severity	Workaround	Check Item
Direct different databases, tables, or partition files to the same path, for example, default warehouse path **/user/hive/warehouse**.	The creation operation may cause disordered data. After a database, table, or partition is deleted, other object data will be lost.	▲ ▲ ▲ ▲ ▲	Do not perform this operation.	Check whether files in the target path are lost.

IoTDB High-Risk Operations

Table 5-16 IoTDB high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Delete data directories.	This operation may cause service information loss.	▲ ▲ ▲	Do not delete data directories manually.	Check whether data directories are normal.
Modify data directory content (file and folder creation).	This operation may cause the IoTDB instance of the node faults.	▲ ▲ ▲	Do not create or modify files or folders in the data directories manually.	Check whether data directories are normal.

Operation	Risk	Severity	Workaround	Check Item
Start or stop basic components independently.	This operation has adverse impact on the basic functions of some services. As a result, service failures occur.	▲ ▲ ▲	Do not start or stop Kerberos, and LDAP basic components independently. Select related services when performing this operation.	Check whether the service status is normal.
Restart or stop services.	This operation may interrupt services.	▲ ▲	Restart or stop services when necessary.	Check whether the service is running properly.

Kafka High-Risk Operations

Table 5-17 Kafka high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Delete Topic	This operation may delete existing topics and data.	▲ ▲ ▲	Kerberos authentication is used to ensure that authenticated users have operation permissions. Ensure that topic names are correct.	Check whether topics are processed properly.
Delete data directories.	This operation may cause service information loss.	▲ ▲ ▲	Do not delete data directories manually.	Check whether data directories are normal.

Operation	Risk	Severity	Workaround	Check Item
Modify data directory content (file and folder creation).	This operation may cause the Broker instance of the node faults.	▲ ▲ ▲	Do not create or modify files or folders in the data directories manually.	Check whether data directories are normal.
Modify the disk auto-adaptation function using the **disk.adapter.enable** parameter.	This operation adjusts the topic data retention period when the disk usage reaches the threshold. Historical data that does not fall within the storage retention may be deleted.	▲ ▲ ▲	If the retention period of some topics cannot be adjusted, add this topic to the value of **disk.adapter.topic.blacklist**.	Observe the data storage period on the Kafka topic monitoring page.
Modify data directory **log.dirs** configuration.	Incorrect operation may cause process faults.	▲ ▲ ▲	Ensure that the added or modified data directories are empty and that the directory permissions are right.	Check whether data directories are normal.
Reduce the capacity of the Kafka cluster.	This operation may cause quantity reduction of backups of some data duplicates of topic. As a result, some topics cannot be accessed.	▲ ▲	Perform backup operation and then reduce the capacity of the Kafka cluster.	Check whether backup nodes where partitions are located are activated to ensure data security.

Operation	Risk	Severity	Workaround	Check Item
Start or stop basic components independently.	This operation has adverse impact on the basic functions of some services. As a result, service failures occur.	▲ ▲ ▲	Do not start or stop ZooKeeper, Kerberos, and LDAP basic components independently. Select related services when performing this operation.	Check whether the service status is normal.
Restart or stop services.	This operation may interrupt services.	▲ ▲	Restart or stop services when necessary.	Check whether the service is running properly.
Modify configuration parameters.	This operation requires service restart for configuration to take effect.	▲ ▲	Modify configuration when necessary.	Check whether the service is running properly.
Delete or modify metadata.	Modifying or deleting Kafka metadata on ZooKeeper may cause the Kafka topic or service unavailability.	▲ ▲ ▲	Do not delete or modify Kafka metadata stored on ZooKeeper.	Check whether the Kafka topics or Kafka service is available.
Delete metadata backup files.	After Kafka metadata backup files are modified and used to restore Kafka metadata, Kafka topics or the Kafka service may be unavailable.	▲ ▲ ▲	Do not delete Kafka metadata backup files.	Check whether the Kafka topics or Kafka service is available.

KrbServer High-Risk Operations

Table 5-18 KrbServer high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify the **KADMIN_PORT** parameter of KrbServer.	After this parameter is modified, if the KrbServer service and its associated services are not restarted in a timely manner, the configuration of KrbClient in the cluster is abnormal and the service running is affected.	▲ ▲ ▲ ▲ ▲	After this parameter is modified, restart the KrbServer service and all its associated services.	None
Modify the **kdc_ports** parameter of KrbServer.	After this parameter is modified, if the KrbServer service and its associated services are not restarted in a timely manner, the configuration of KrbClient in the cluster is abnormal and the service running is affected.	▲ ▲ ▲ ▲ ▲	After this parameter is modified, restart the KrbServer service and all its associated services.	None
Modify the **KPASSWD_PORT** parameter of KrbServer.	After this parameter is modified, if the KrbServer service and its associated services are not restarted in a timely manner, the configuration of KrbClient in the cluster is abnormal and the service running is affected.	▲ ▲ ▲ ▲ ▲	After this parameter is modified, restart the KrbServer service and all its associated services.	None

Operation	Risk	Severity	Workaround	Check Item
Modify the domain name of Manager system.	After the domain name is modified, if the KrbServer service and its associated services are not restarted in a timely manner, the configuration of KrbClient in the cluster is abnormal and the service running is affected.	▲ ▲ ▲ ▲ ▲	After this parameter is modified, restart the KrbServer service and all its associated services.	None
Configure cross-cluster mutual trust relationships.	This operation will restart the KrbServer service and all associated services, affecting the management and maintenance and services of the cluster.	▲ ▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.

LdapServer High-Risk Operations

Table 5-19 LdapServer high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify the **LDAP_SERVER_PORT** parameter of LdapServer.	After this parameter is modified, if the LdapServer service and its associated services are not restarted in a timely manner, the configuration of LdapClient in the cluster is abnormal and the service running is affected.	▲ ▲ ▲ ▲ ▲	After this parameter is modified, restart the LdapServer service and all its associated services.	None

Operation	Risk	Severity	Workaround	Check Item
Restore LdapServer data.	This operation will restart Manager and the entire cluster, affecting management, maintenance, and services of the cluster.	▲ ▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Replace the Node where LdapServer is located.	This operation will interrupt services deployed on the node. If the node is a management node, the operation will restart all OMS processes, affecting the cluster management and maintenance.	▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check for uncleared alarms, and check whether the cluster management and maintenance and services are normal.
Change the password of LdapServer.	The LdapServer and Kerberos services need to be restarted during the password change, affecting the management, maintenance, and services of the cluster.	▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	None

Operation	Risk	Severity	Workaround	Check Item
Restart the node where LdapServer is located.	Restarting the node without stopping the LdapServer service may cause LdapServer data damage.	▲ ▲ ▲ ▲ ▲	Restore LdapServer using LdapServer backup data	None

Loader High-Risk Operations

Table 5-20 Loader high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Change the floating IP address of a Loader instance (**loader.float.ip**).	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether the Loader UI can be connected properly.
Modify the Loader instance start parameter **LOADER_GC_OPTS**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Clear table contents when adding data to HBase.	This operation will clear original data in the target table.	▲ ▲	Ensure that the contents in the target table can be cleared before the operation.	Check whether the contents in the target table can be cleared before the operation.

MemArtsCC High-Risk Operations

Table 5-21 MemArtsCC high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Delete or modify the ZooKeeper key node. (The node is specified by configuration item **zk_key_node**. The default value is **/memartscc/key**.)	This operation will prevent MemArtsCC from reading the cache, which will increase the amount of traffic read from OBS.	▲ ▲ ▲	Do not manually delete or modify the key node.	Check whether the cache is normal based on the monitoring metrics.
Delete or modify the ZooKeeper root node. (The root node is specified by the **zk_root_node** configuration item. The default value is **/memartscc**.)	This operation will lead to MemArtsCC service exceptions.	▲ ▲ ▲	Do not manually delete or modify the root node.	Check whether the service is normal.

Operation	Risk	Severity	Workaround	Check Item
Modify the permission on the cache path file.	This operation will report an alarm indicating that the disk is unavailable and affect the cache.	▲ ▲ ▲	Do not manually modify the permission on the cache path file.	Check whether the cache is normal and whether an alarm indicating that the disk is unavailable is reported based on the related monitoring metrics.
Start or stop basic components.	This operation will lead to MemArtsCC service exceptions.	▲ ▲	Do not start or stop ZooKeeper basic components independently. Select related services when performing this operation.	Check whether the service status is normal.
Restart or stop services.	This operation may prevent MemArtsCC from reading the cache, which will increase the amount of traffic read from OBS.	▲ ▲	Restart or stop services when necessary.	Check whether the service status is normal.

MOTService High-Risk Operations

Table 5-22 MOTService high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Change the MOTService password.	The services need to be restarted for the password change to take effect. The services are unavailable during the restart. After the password is changed, the service configuration needs to be updated accordingly.	▲ ▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check whether services are restored.
Restore MOTService data.	After the data is restored, the data generated after the data backup and before the data restoration is lost.	▲ ▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check whether services are restored.

Operation	Risk	Severity	Workaround	Check Item
Delete a MOTService instance.	Data may be lost and cannot be restored, service running may be faulty, or other instance configurations may expire.	▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time. **NOTE** This function applies only to MRS physical machine clusters.	Check whether services are restored.

Operation	Risk	Severity	Workaround	Check Item
Perform active/standby MOTService switchover.	MOTService is unavailable during the switchover.	▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time. Before performing an active/standby switchover, ensure that data synchronization between the active and standby nodes is complete. Otherwise, data may be lost. If the "ALM-46004 Data Inconsistency Between Active and Standby MOTService Nodes" alarm is generated, clear the alarm before proceeding with subsequent operations.	Check whether the active/standby switchover is complete and whether services are restored.

Operation	Risk	Severity	Workaround	Check Item
Change the MOTService floating IP address.	The MOTService service needs to be restarted for the configuration to take effect. The MOTService service is inaccessible during the restart. After the modification, you need to update the BLU configuration in Containers to ensure that the modification takes effect. Before the modification, ensure that the new floating IP address is not in use. If it is in use, the configuration fails and MOTService fails to start.	▲ ▲ ▲ ▲ ▲	Ensure that this operation is necessary.	Check whether the service can be started properly and whether the service is restored.
Stop or restart the MOTService service.	This operation will interrupt services. (The database memory needs to be loaded during the restart, which takes a long time. You are advised to perform a rolling restart.)	▲ ▲	Ensure that this operation is necessary.	Check whether services are restored.
Restart the active MOTService instance.	Services are interrupted.	▲ ▲	Ensure that this operation is necessary.	Check whether services are restored.
Do not replicate large table data. For example: insert into *table1* select * from *tables2*;	Data replication between tables with large volumes of data occupies too much memory. As a result, the service becomes abnormal.	▲ ▲ ▲ ▲ ▲	Use gs_dump to perform logical backup.	Check whether the data is correctly backed up.

Ranger High-Risk Operations

Table 5-23 Ranger high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify the Ranger instance **GC_OPTS** startup parameters.	This operation may cause Ranger instance start failures.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether the service is running properly.
Delete a component service on the Ranger page.	This operation may cause service permission exceptions.	▲ ▲	Do not perform this operation.	Check whether related component service operations are normal.

Redis High-Risk Operations

Table 5-24 Redis high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Change the memory size.	If Redis occupies too much memory, the process may be stopped by the OS or OOM may occur.	▲ ▲ ▲ ▲	Properly plan the maximum memory size of each process based on the OS memory to prevent OOM.	Check whether the service is running properly.
Restart services or instances.	This operation may interrupt services.	▲ ▲	Restart or stop services when necessary.	Check whether the service is running properly and whether interrupted operations are restored.

Operation	Risk	Severity	Workaround	Check Item
Synchronize configurations (by restarting the required service).	During the restart, the Redis cluster where the instance is located cannot provide services.	▲ ▲	Restart or stop services when necessary.	Check whether the service is running properly and whether interrupted operations are restored.
Stop services or instances.	This operation may interrupt services.	▲ ▲	Restart or stop services when necessary.	Check whether the services are properly stopped.
Delete services or instances.	Data stored in Redis is lost, and Redis cannot provide services.	▲ ▲ ▲	Do not perform this operation unless data stored in Redis needs to be discarded. **NOTE** This function applies only to MRS physical machine clusters.	Check whether the service is successfully deleted.
Scale in or out a cluster.	Read and write of some service data may become faulty because data is migrated.	▲ ▲ ▲	Minimize the Redis cluster load or ensure that the Redis cluster is not providing services as possible.	Check whether cluster scale-in or scale-out is successful.
Change the password.	During password change, Redis cannot provide services.	▲ ▲	Change the password when necessary.	Use the new password for login again and check whether the login is successful.

Operation	Risk	Severity	Workaround	Check Item
Delete a Redis cluster.	Data stored in the Redis cluster will be lost, and services that access the cluster will be interrupted.	▲ ▲ ▲	Do not perform this operation unless data stored in Redis needs to be discarded.	Check whether the Redis cluster is deleted.
Delete or modify backup files.	The Redis process will be restarted or become faulty, and data will be lost.	▲ ▲ ▲ ▲	Do not perform this operation.	Check whether the Redis cluster is normal and whether data is lost.
Delete or modify metadata.	The Redis service may be faulty.	▲ ▲ ▲ ▲	Do not perform this operation.	Check whether the Redis service is normal.
Modify file permissions.	The Redis service may be faulty.	▲ ▲ ▲ ▲	Do not perform this operation.	Check whether related service operations are normal.
Delete or modify files.	The Redis service may be faulty.	▲ ▲ ▲ ▲	Do not perform this operation.	Check whether the Redis service is normal.

RTDService High-Risk Operations

Table 5-25 RTDService high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify an RTDService tenant.	Adding or deleting a database may cause service access problems. For example: • Changing the floating IP address of an existing database is similar to replacing the database. If the new database does not contain data such as tables and stored procedures corresponding to the service, the service reports an error. • Change the port number. If the port number is changed incorrectly, service access may fail. • After the floating IP address of the Farmer cluster is changed, BLU will be deployed on the new Farmer cluster. In this case, the BLU operation on the old Farmer cluster will fail. • After the Farmer cluster password is changed, if the new password is different from the password of the cluster user, the user will be connected for multiple times. As a result, the user is locked.	▲ ▲ ▲ ▲ ▲	Before the modification, confirm the necessity of the operation, provide a modification scheme, and evaluate the impact. Back up data before the modification to ensure that a rollback can be performed in time when a fault occurs.	Check whether service operations are affected. Check whether data synchronization is affected. Check whether service computing is affected.

Operation	Risk	Severity	Workaround	Check Item
Delete an RTDService instance.	This operation may cause service running faults or other instance configurations to expire.	▲ ▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time. **NOTE** This function applies only to MRS physical machine clusters.	Check whether there are uncleared alarms and whether the cluster management and maintenance are normal.
Perform active/standby RTDService switchover.	During the RTDService switchover, RTDService is unavailable.	▲ ▲	Before performing the operation, ensure that the operation is necessary, and that no other management and maintenance operations are performed at the same time.	Check whether the active/standby switchover is complete and whether services are restored.

Operation	Risk	Severity	Workaround	Check Item
Change the RTDService floating IP address.	The RTDService service needs to be restarted for the configuration to take effect. The RTDService service is inaccessible during the restart. Before the modification, ensure that the new floating IP address is not in use. If it is in use, the configuration fails and RTDService fails to start.	▲ ▲ ▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.

Solr High-Risk Operations

Table 5-26 Solr high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify Solr instance port parameters (SOLR_PORT, SOLR_SEC_PORT, and SOLR_CONTROL_PORT).	Maloperation will cause instance start and stop abnormality.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether service instances can be started or stopped properly.

Operation	Risk	Severity	Workaround	Check Item
Modify the Solr parameter **INDEX_STORED_ON_HDFS**.	• If the configuration of the Collection in the configuration set **solrconfig.xml** is **<directoryFactory name=""DirectoryFactory""class=""${solr.directoryFactory:solr.NRTCachingDirectoryFactory}"">**, when the **INDEX_STORED_ON_HDFS** parameter is modified, the index storage location of the Collection that adopts the configuration changes, and the Collection needs to be indexed again. Index data in the original storage location will not be deleted automatically. • If the configuration of the Collection in the configuration set **solrconfig.xml** is **<directoryFactory name=""DirectoryFactory""class=""solr.NRTCachingDirectory-Factory"">**, the index of the Collection that adopts the configuration will not be affected by the modification of the **INDEX_STORED_ON_HDFS** parameter.	▲ ▲ ▲	Identify the Collection that will be affected when the parameter is modified. • To avoid the effect caused by the parameter, change **<directoryFactory name=""DirectoryFactory""class=""${solr.directoryFactory:solr.NRTCachingDirectoryFactory}"">** in the configuration set **solrconfig.xml** of the Collection to **<directoryFactoryname=""DirectoryFactory"" class=""solr.NRTCachingDirectoryFactory"">**. • Index the Collection that is affected again.	None

Operation	Risk	Severity	Workaround	Check Item
Delete two SolrServerAdmin instances at the same time.	The operation may cause unavailable services.	▲ ▲ ▲	Normally, two SolrServerAdmins are required. You can delete only one of them. After the deletion, restore the original node to two instances as soon as possible. To migrate instances, migrate two instances at the same time and migrate data in advance. For details, see *MapReduce Service (MRS) 3.3.1-LTS Physical Machine Cluster Component Migration Guide (for Huawei Cloud Stack 8.3.1)*. **NOTE** This function of implement instance migration applies only to the physical machine cluster.	Check whether the service status is normal.

Operation	Risk	Severity	Workaround	Check Item
Modify the startup parameters **SOLR_GC_OPTS** and **SOLR_HEAPSIZE** of a Solr instance.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid. SOLR_GC_OPTS does not conflict with SOLR_HEAPSIZE.	Check whether services can be started properly.

Spark High-Risk Operations

Table 5-27 Spark high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Modify the configuration item **spark.yarn.queue**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.

Operation	Risk	Severity	Workaround	Check Item
Modify the configuration item **spark.driver.extraJavaOptions**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Modify the configuration item **spark.yarn.cluster.driver.extraJavaOptions**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Modify the configuration item **spark.eventLog.dir**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Modify the configuration item **SPARK_DAEMON_JAVA_OPTS**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.

Operation	Risk	Severity	Workaround	Check Item
Delete all JobHistory instances.	The event logs of historical applications are lost.	▲ ▲	Reserve at least one JobHistory instance.	Check whether historical application information is included in JobHistory.
Delete or modify the **/user/spark/jars/8.3.1/spark-archive.zip** file in HDFS.	JDBCServer fails to start and service functions are abnormal.	▲ ▲ ▲	Delete **/user/spark/jars/8.3.1/spark-archive.zip** and wait for 10 to 15 minutes until the .zip package is automatically restored.	Check whether services can be started properly.

Yarn High-Risk Operations

Table 5-28 Yarn high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Delete or change data directories **yarn.nodemanager.local-dirs** and **yarn.nodemanager.log-dirs**	This operation may cause service information loss.	▲ ▲ ▲	Do not delete data directories manually.	Check whether data directories are normal.

ZooKeeper High-Risk Operations

Table 5-29 ZooKeeper high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Delete or change ZooKeeper data directories.	This operation may cause service information loss.	▲ ▲ ▲	Follow the capacity expansion guide to change the ZooKeeper data directories.	Check whether services and associated components are started properly.
Modify the ZooKeeper instance start parameter **GC_OPTS**.	Services cannot start properly.	▲ ▲	Strictly follow the prompt information when modifying related configuration items. Ensure that new values are valid.	Check whether services can be started properly.
Modify the znode ACL information in ZooKeeper.	If znode permission is modified in ZooKeeper, other users may have no permission to access the znode and some system functions are abnormal.	▲ ▲ ▲ ▲	During the modification, strictly follow the ZooKeeper Configuration Guide and ensure that other components can use ZooKeeper properly after ACL information modification.	Check that other components that depend on ZooKeeper can properly start and provide services.

Elasticsearch High-Risk Operations

Table 5-30 Elasticsearch high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Change a non-security mode to the security mode.	Some indexes cannot be accessed in security mode due to misoperations.	▲ ▲ ▲ ▲	Perform operations by strictly following the instructions provided in "Switching the Elasticsearch Security Mode" in *MapReduce Service (MRS) 3.3.1-LTS User Guide (for Huawei Cloud Stack 8.3.1) in the MapReduce Service (MRS) 3.3.1-LTS Usage Guide (for Huawei Cloud Stack 8.3.1)*.	Check whether the service is running properly.

Operation	Risk	Severity	Workaround	Check Item
Modify the data directory parameter **elasticsearch.data.path**.	This operation will cause service data loss in the original data directory.	▲ ▲ ▲ ▲	Perform operations by strictly following the instructions provided in "Customizing a Data Catalog" in *MapReduce Service (MRS) 3.3.1-LTS User Guide (for Huawei Cloud Stack 8.3.1) in the MapReduce Service (MRS) 3.3.1-LTS Usage Guide (for Huawei Cloud Stack 8.3.1)*.	Check whether services can be started properly.
Delete an index.	Incorrect operations may lead to service data loss.	▲ ▲ ▲	Ensure that indexes are deleted only when necessary. Only invalid and expired indexes can be deleted. Indexes that are being used cannot be deleted.	Check whether the service is running properly.
Change the number of index replicas to 0.	This operation will reduce service data reliability.	▲ ▲	Do not manually change the number of index replicas to 0.	Check whether the service and indexes are normal.

Operation	Risk	Severity	Workaround	Check Item
Delete more than half of the EsMaster instances at the same time.	This operation will lead to service exceptions.	▲ ▲ ▲ ▲	Perform operations by strictly following the instructions provided in "Reducing Elasticsearch Capacity" in *MapReduce Service (MRS) 3.3.1-LTS Physical Machine Cluster Capacity Expansion Guide (for Huawei Cloud Stack 8.3.1)*. **NOTE** This function applies only to MRS physical machine clusters.	Check whether the service status is normal.
Set **scroll_id** to **_all** when the **Clear scroll API** is used.	This operation clears all scroll query caches, causing exceptions in other ongoing scroll query services.	▲ ▲ ▲	When the **Clear scroll API** is used, set **scroll_id** to the **scroll** of the scroll query service.	Check whether the service is running properly.

Doris High-Risk Operations

Table 5-31 Doris high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Delete data directories.	This operation may cause service information loss.	▲ ▲ ▲	Do not delete data directories manually.	Check whether data directories are normal.
Modify data directory content (file/ folder creation).	This operation may cause the Doris instance faults on the node.	▲ ▲ ▲	Do not create or modify files or folders in the data directories manually.	Check whether data directories are normal.
Restart/ Stop the Doris service.	This operation may interrupt services.	▲ ▲	Restart or stop services when necessary.	Check whether the service is running properly.
Modify the files or configuration files in the Doris installation directory.	The operation may cause unavailable services.	▲ ▲	Do not modify the files in the Doris installation directory.	Check whether the service is running properly.
Delete more than **n/2 + 1** FE instances.	The service is unavailable due to this operation.	▲ ▲ ▲	Do not delete FE instances greater than or equal to **n/2 + 1** (**n** indicates the number of FE instance nodes).	Check whether the service is running properly.

LakeSearch High-risk Operations

Table 5-32 LakeSearch high-risk operations

Operation	Risk	Severity	Workaround	Check Item
Deleting a data directory	This operation may cause service information loss.	▲ ▲ ▲	Do not delete data directories manually.	Check whether data directories are normal.
Restarting or stopping Elasticsearch, DBService, and HBase	This operation may interrupt services.	▲ ▲ ▲	Do not perform this operation unless necessary.	Check whether these services are normal.
Restarting or stopping the LakeSearch Service	This operation may interrupt services.	▲ ▲	Restart or stop services when necessary.	Check whether the service is running properly.
Modifying the files or configuration files in the LakeSearch installation directory	The service may become unavailable.	▲ ▲	Do not modify files in the LakeSearch installation directory unless necessary.	Check whether the service is running properly.
Delete more than **n/2 + 1** LakeSearch instances.	The service may become unavailable.	▲ ▲ ▲	Do not delete LakeSearch instances greater than or equal to **n/2 + 1** (**n** indicates the number of LakeSearch instance nodes).	Check whether the service is running properly.

6 Common Operations

6.1 Logging In to Common Portals

Logging In to ManageOne Maintenance Portal

Step 1 Log in to ManageOne Maintenance Portal as a system administrator using a browser.

- URL: **https://***Domain name of ManageOne Maintenance Portal*:**31943**. Alternatively, access **https://***Address for accessing the ManageOne unified portal* to log in to the ManageOne main portal and choose OperationCenter to access ManageOne Maintenance Portal.

  Obtain the domain name of ManageOne Maintenance Portal on the "2.3 Portal" sheet of the deployment parameter table exported from HCC Turnkey during the installation of Huawei Cloud Stack basic services.

- Login using a password: Enter the username and password.

  Default account: bss_admin. To obtain the default password, contact the system administrator or refer to the default password of the ManageOne Maintenance Portal account on the "Type A (Portal)" sheet in ***Huawei Cloud Stack 8.3.1 Account List*** .

  NOTE

  For ManageOne upgraded from 8.2.0 or earlier, the preset username is **admin**.

- Login using a USB key: Insert a USB key with preset user certificates, select the required device and certificate, and enter a PIN.

Step 2 Click **Log In**.

----End

Logging In to ManageOne Operation Portal (VDC administrator)

Step 1 Log in to ManageOne as the VDC administrator using a browser.

- URL in non-B2B scenarios: **https://***Domain name of ManageOne Operation Portal*, for example, **https://console.demo.com**

  URL in B2B scenarios: **https://***Domain name of ManageOne Tenant Portal*, for example, **https://tenant.demo.com**

You can log in using a password or a USB key.

URL of the unified portal: **https://***Address for accessing the ManageOne unified portal*, for example, **https://console.demo.com/moserviceaccesswebsite/unifyportal#/home**. On the home page, choose **Self-service Cloud Service Center**.

- Login using a password: Enter the username and password.

  Enter the username and password of the VDC administrator .
- Login using a USB key: Insert a USB key with requested user certificates, select the required device and certificate, and enter a PIN.

Step 2 Select your region and project from the drop-down list on the top menu bar.

----End

Logging In to Service OM

Step 1 Log in to ManageOne Maintenance Portal as a system administrator using a browser.

Step 2 On the home page, click Service_OM and select the target region to switch to the Service OM page.

----End

Logging In to CloudScope

Step 1 Log in to CloudScope using a browser as a system administrator.

- URL: **https://***CloudScope domain name*, for example, **https://cloudscope.demo.com**

  For details about the URL for accessing CloudScope, see the COP information on the "Portal" sheet of deployment parameter table exported from HCC Turnkey during Auto Change Platform installation.
- For details about the default account information, see ***Huawei Cloud Stack 8.3.1 Account List*** .

----End

6.2 Checking the IP Address of the Cloud Service Management VM

Step 1 Use a browser to log in to ManageOne Maintenance Portal as a system administrator and go to the **Service OM** page of the region where the cloud service is located.

Step 2 Choose **Services** > **Resource** > **Compute Resource** from the main menu.

Step 3 Click the **VMs** tab, enter a keyword in the search box to search for the VM name, for example, **MRS_DB**, and record the IP address of the VM.

Step 4 Log in to the corresponding node as the user of the VM.

For details about the default user information of each service VM, see ***Huawei Cloud Stack 8.3.1 Account List*** .

----End

6.3 Checking the MRS Container Status

Step 1 Log in to the **EICommon-Region-Master-01** VM as user **opsadmin**. For details, see **Logging In to an MRS Management Node**.

su - root

Obtain the default password by referring to ***Huawei Cloud Stack 8.3.1 Account List*** or contacting the system administrator.

Step 2 Run the following command to query the names of all MRS containers in the cluster and check information of the containers:

kubectl get pods -n mrs -o wide

```
NAME                          READY   STATUS    RESTARTS   AGE     IP           NODE           NOMINATED
NODE   READINESS GATES
mrsapigw-6f56bc476d-c5cs9     1/1     Running   0          19h     10.16.0.69   10.69.26.187   <none>
<none>
mrsapigw-6f56bc476d-smqn8     1/1     Running   0          5d13h   10.16.0.20   10.69.26.197
<none>         <none>
mrsdeployer-5988d78867-2prk6  1/1     Running   0          43h     10.16.0.55   10.69.26.194   <none>
<none>
mrsdeployer-5988d78867-8lb5p  1/1     Running   0          43h     10.16.0.25   10.69.26.197   <none>
<none>
mrsdeployer-5988d78867-gwqb7  1/1     Running   0          43h     10.16.0.86   10.69.26.189
<none>         <none>
```

Step 3 Run the **kubectl exec -ti -n mrs** *Container name* **bash** command to access the VM where the container is located.

For example, run the **kubectl exec -ti -n mrs** *mrsapigw-6f56bc476d-pxcvk* **bash** command to access the container node.

----End

6.4 Determining the Active/Standby Status of MRS-DB Nodes

Symptom

MRS-DB nodes MRS_DB-01 and MRS_DB-02 are active and standby nodes. If an active DB node is faulty, the standby DB node automatically becomes the active one, ensuring system reliability.

Procedure

Step 1 Log in to the MRS_DB-01 VM as user **opsadmin**. For details, see section **Logging In to an MRS Management Node**. Then, run the following command to switch to user **root**:

su - root

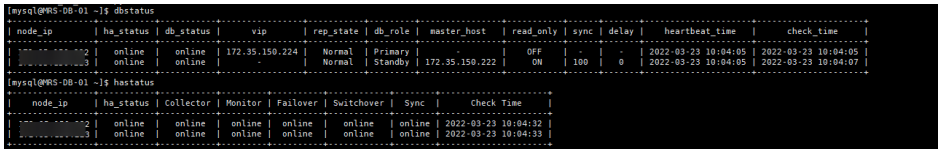
See ***Huawei Cloud Stack 8.3.1 Account List*** or contact the system administrator to obtain the default password.

Step 2 Run the following commands to query the primary/standby status of the database and the floating IP address (vip) of the database:

su - mysql

dbstatus

hastatus



----End

6.5 Logging In to an MRS Management Node

Huawei Cloud Stack provides a security O&M channel through the CLI terminal (an optional module). When you install the security O&M channel, the way of logging in to backend nodes changes. This section describes how to log in to a backend node in different scenarios.

The module provides security O&M capabilities for ManageOne Maintenance Portal, including SSH login without password, command execution or interception, file upload and download, and O&M command audit. The module provides a unified O&M portal. This enhances O&M management, control, and audit, simplifies the remote O&M access process, improves O&M efficiency, and enhances account and password security.

The module is selected by default and can be deselected. It is not upgraded by default and can be added after the upgrade.

NOTE

The CLI supports the following accounts in different scenarios:

- Accounts of cloud services where the security O&M channel is not installed can only be used to log in to nodes over SSH.
- Accounts of cloud services where the security O&M channel is installed but that are not reclaimed can only be used to log in to nodes over SSH.
- Accounts of cloud services where the secure O&M channel is installed and that are reclaimed can be used to log in to nodes over SSH by applying for a password because its password has been changed to a random one after reclamation.
- Accounts of cloud services where the secure O&M channel is installed, that are reclaimed, and that have obtained the one-click login permission can be used to log in to nodes through the CLI terminal or over SSH by applying for a password.

Prerequisites

- You have logged in to ManageOne Maintenance Portal as a system administrator.
- You have obtained the permission to perform operations on the CLI.

- If the command whitelist is enabled, only whitelisted commands can be executed on the CLI. If the function is disabled, any commands can be executed on the CLI.

Logging In to an MRS Management Node Through the CLI Terminal

Step 1 Check whether secure O&M channel has been installed.

Choose **O&M** from the main menu and check whether the CLI terminal is displayed in the navigation pane.

- If yes, the secure O&M channel has been installed. Go to **Step 2**.
- If no, the secure O&M channel is not installed. In this case, log in to a management node over SSH. For details, see **Logging In to an MRS Management Node Over SSH**.

Step 2 Check whether the account has been managed by an external system.

1. Choose **O&M** > **Accounts** from the main menu.
2. Check whether the account for logging in to the backend node has been remotely managed based on information displayed in the **Application/Cloud Service**, **Resource Name**, and **Account Management Party** columns.
   - If yes, use an SSH client, such as PuTTY, to log in to a node.
   - If no, go to **Step 3**.

Step 3 Use either of the following methods to log in to a backend node:

- Login using SSH

  Apply for permission to use an account password to log in to the node over SSH. For details, see section "Account Request" > "Creating a Request for Obtaining Passwords" in the **Huawei Cloud Stack 8.3.1 O&M Guide**.
- One-click login on the CLI
  a. Choose **O&M** > **CLI** from the main menu.
  b. In the navigation pane on the left, expand the corresponding nodes one by one and check whether the account for logging in to the backend node is displayed.
     - If yes, the account has obtained the permission for one-click login. Go to **Step 4**.
     - If no, apply for the one-click login permission to log in to the node. For details, see section "Account Request" > "Creating a Request for Obtaining Passwords" in the **Huawei Cloud Stack 8.3.1 O&M Guide**. Then, go to **Step 4**.

Step 4 Click **Execute Command** to log in to the backend node. Then, proceed with other operations.

NOTE

Only commands in the trustlist can be executed. If the following error message is displayed during the execution, contact the administrator to add the commands to the trustlist. For details, see the **Huawei Cloud Stack 8.3.1 O&M Guide**.
Your command is highly risky. Please check the whitelisted commands.

----End

Logging In to an MRS Management Node Over SSH

Step 1 Log in to ManageOne Maintenance Portal as the system administrator. In the **Common Links** navigation tree, click Service_OM and select a region to go to the Service OM page.

Step 2 Choose **Services** > **Resource** > **Compute Resource**.

Step 3 Click the **VMs** tab, enter a keyword in the search box to search for the VM name, for example, **EICommon-Region-Master**, **MRS_DB**, **Console-DB**, or **Console-Static**, and record the IP address of the VM.

Step 4 Use PuTTY to log in to any VM whose **Status** is **Running** as user **opsadmin**. For example, log in to the **EICommon-Region-Master-01**, **MRS_DB-01**, **Console-DB-01**, or **Console-Static-01** VM.

Obtain the default password by referring to ***Huawei Cloud Stack 8.3.1 Account List*** or contacting the system administrator.

NOTE

If **Status** of all VMs is not **Running**, contact the cluster administrator.

----End

