MapReduce Service

Getting Started

Issue	01
Date	2022-09-14

HUAWEI

HUAWEI TECHNOLOGIES CO., LTD.

Copyright © Huawei Technologies Co., Ltd. 2022. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

Trademarks and Permissions


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

Notice

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

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

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: https://www.huawei.com

Email: support@huawei.com

Contents

 Copyright © Huawei Technologies Co., Ltd.

1 Buying and Using an MRS Cluster

1.1 How to Use MRS

MapReducce Service is a Huawei Cloud service that is used to deploy and manage Hadoop clusters. MRS provides enterprise-class big data clusters on the cloud. Tenants can fully control these clusters and easily run big data components such as Hadoop, Spark, HBase, and Kafka in them.

MRS is easy to use. You can execute various tasks and process or store PB-level data using computers connected in a cluster.

The procedure of using MRS is as follows:

1. On the MRS console, purchase clusters and specify these clusters for offline data analysis and stream processing, and specify the Elastic Cloud Server (ECS) instance specifications, quantity, data disk types (common I/O, high I/O, or ultra-high I/O), as well as components to be installed in the clusters.
2. Develop a data processing program. For details about how to quickly develop such a program and execute it properly, see the sample code and tutorials provided in **Method of Building an MRS Sample Project**.
3. Upload the prepared program and data files to Object Storage Service (OBS) or the HDFS in the cluster.
4. After a cluster is created, you can directly add jobs and run your programs or SQL statements to process and analyze data.
5. MRS provides you with MRS Manager, an enterprise-class unified management platform of big data clusters, helping you quickly know the health status of services and hosts. Through graphical metric monitoring and customization, you can obtain critical system information in a timely manner. In addition, you can modify service attribute configurations based on service performance requirements, and start or stop clusters, services, and role instances in one click.
6. Terminate the cluster if it is no longer needed after job execution. The terminated cluster is no longer billed.

 Copyright © Huawei Technologies Co., Ltd.

1.2 Creating a Cluster

The first step of using MRS is to purchase a cluster. This section describes how to create a cluster on the MRS console.

Procedure

Step 1 Log in to the MRS console.

Step 2 Click **buy clusters** to access the **buy clusters** page.

NOTE

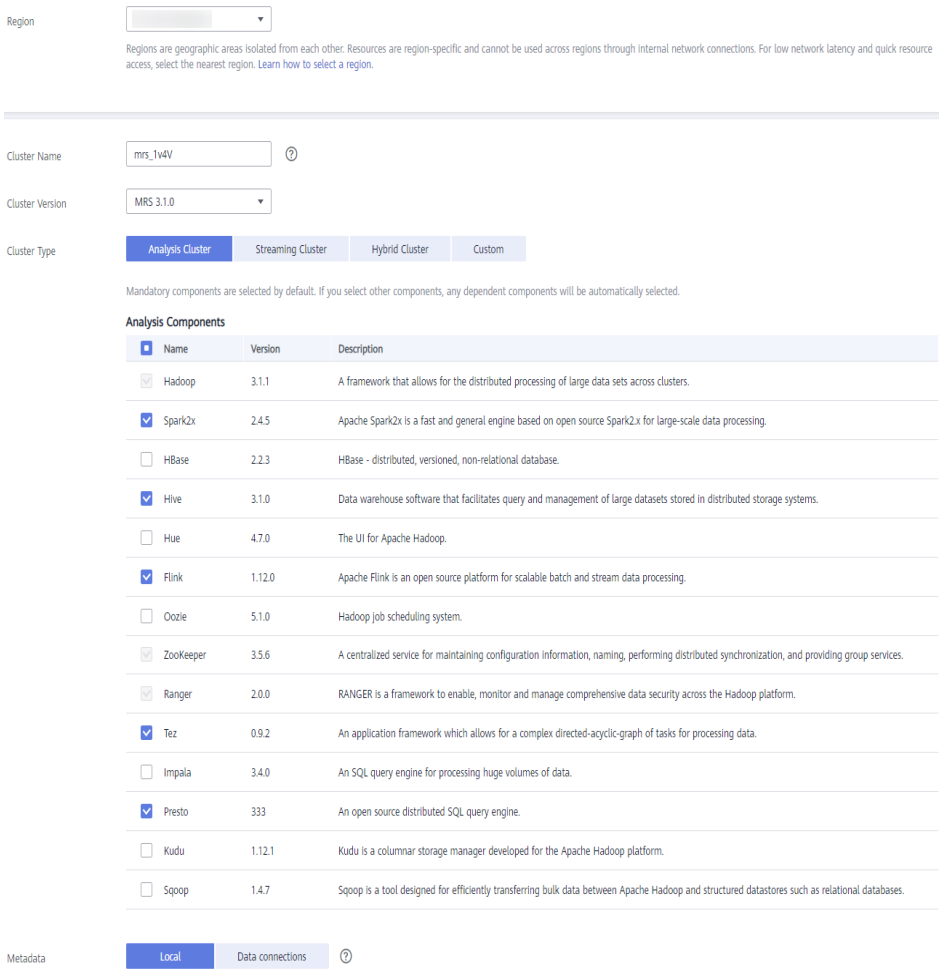
When creating a cluster, pay attention to quota notification. If a resource quota is insufficient, increase the resource quota as prompted and create a cluster.

Step 3 On the page for purchase a cluster, click the **Custom Config** tab.

Step 4 Configure cluster software information.

- **Region**: Use the default value.
- **Cluster Name**: You can use the default name. However, you are advised to include a project name abbreviation or date for consolidated memory and easy distinguishing, for example, **mrs_20180321**.
- **Cluster Version**: Select the latest version, which is the default value.
- **Cluster Type**: Use the default **Analysis Cluster**.
- **Component**: Select components such as Spark2x, HBase, and Hive for the analysis cluster. For a streaming cluster, select components such as Kafka and Storm. For a hybrid cluster, you can select the components of the analysis cluster and streaming cluster based on service requirements.
- **Metadata**: Retain the default value.

 Copyright © Huawei Technologies Co., Ltd.



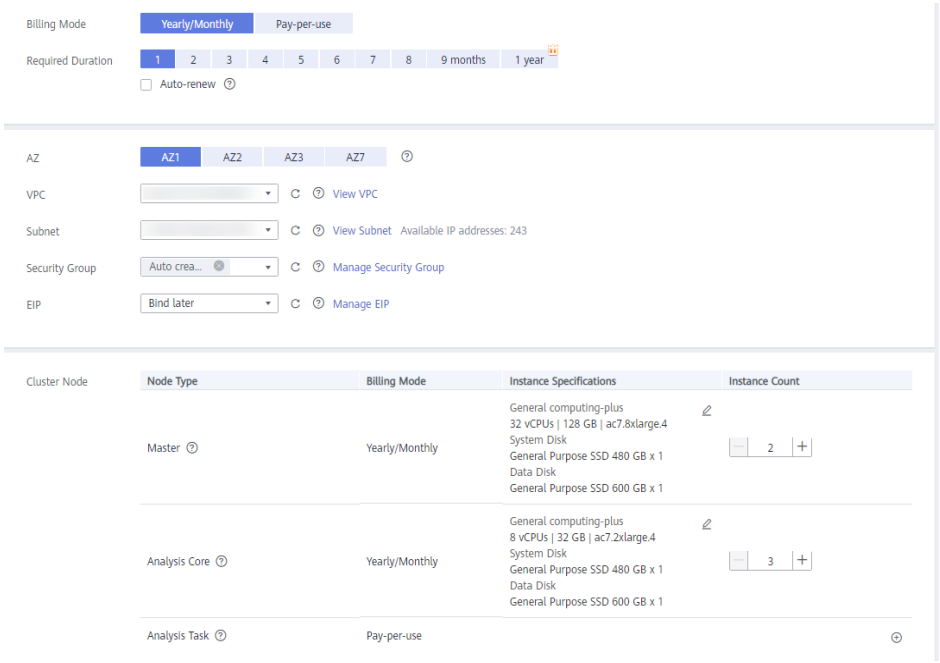
NOTE

For versions earlier than MRS 3.x, select components such as Spark, HBase, and Hive for the analysis cluster.

Step 5 Click **Next**.

- **Billing Mode**: Use the default value.
- **AZ**: Use the default value.
- **VPC**: Use the default value. If there is no available VPC, click **View VPC** to access the VPC console and create a new VPC.
- **Subnet**: Use the default value.
- **Security Group**: Select **Auto create**.
- **EIP**: Select **Bind later**.
- **Instance Specifications**: Retain the default settings for master and core nodes or select proper specifications based on service requirements.
- **System Disk**: Select **General Purpose SSD** and retain the default space.
- **Data Disk**: Select **General Purpose SSD** and retain the default space.
- **Instance Count**: The default number of Master nodes is 2, and that of Core nodes is 3.

 Copyright © Huawei Technologies Co., Ltd.



Step 6 Click **Next**. The **Set Advanced Options** page is displayed. Configure the following parameters. Retain the default settings for the other parameters.

- Kerberos authentication:
  - **Kerberos Authentication**: Disable Kerberos authentication.
  - **Username**: name of the Manager administrator. **admin** is used by default.
  - **Password**: password of the Manager administrator.
- **Login Mode**: Select a mode for logging in to an ECS.
  - **Password**: Set a password for logging in to an ECS.
  - **Key Pair**: Select a key pair from the drop-down list. Select **"I acknowledge that I have obtained private key file *SSHkey-xxx* and that without this file I will not be able to log in to my ECS.**" If you have never created a key pair, click **View Key Pair** to create or import a key pair. And then, obtain a private key file.
- **Secure Communications**: Select **Enable**.

Step 7 Click **Buy Now**.

If Kerberos authentication is enabled for a cluster, check whether Kerberos authentication is required. If yes, click **Continue**. If no, click **Back** to disable Kerberos authentication and then create a cluster.

Step 8 Click **Back to Cluster List** to view the cluster status.

It takes some time to create a cluster. The initial status of the cluster is **Starting**. After the cluster has been created successfully, the cluster status becomes **Running**.

----End

 Copyright © Huawei Technologies Co., Ltd.

1.3 Uploading Data

On the **Files** page, you can create and delete HDFS directories, as well as import, export, and delete files in an analysis cluster.

For clusters with Kerberos authentication enabled, synchronize IAM users before performing operations on the **Files** page. On the cluster details page, click **Dashboard** and click **Synchronize** on the right of **IAM User Sync** to synchronize IAM users.

Background

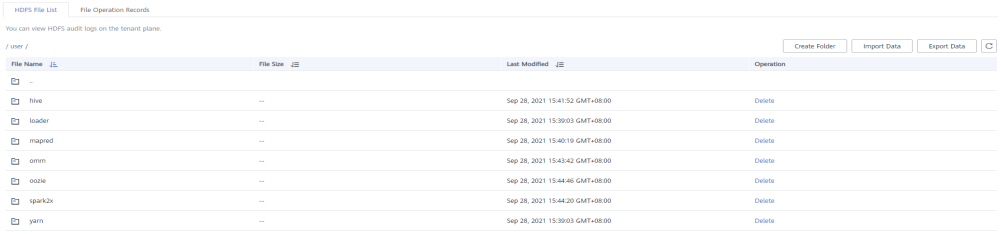
MRS clusters generally process data from OBS or HDFS. OBS provides you with the data storage capabilities that are massive, secure, reliable, and cost-effective. MRS can directly process data in OBS. You can browse, manage, and use data both on the management console and on the OBS Client. If you need to import OBS data into the HDFS system of the cluster for processing, perform the steps in this section.

Importing Data

Currently, MRS can import data from OBS to the HDFS. The file upload rate decreases with the increase of the file size. This mode applies to scenarios where the data volume is small.

You can perform the following steps to import files and directories:

1. Log in to the MRS console.
2. Choose **Clusters** > **Active Clusters**, and click the name of the target cluster to enter the cluster details page.
3. Click **Files** to go to the file management page.
4. Select **HDFS File List**.



5. Go to the data storage directory, for example, **bd_app1**.

   The **bd_app1** directory is only an example. You can use any directory on the page or create a new one.

   The requirements for creating a folder are as follows:

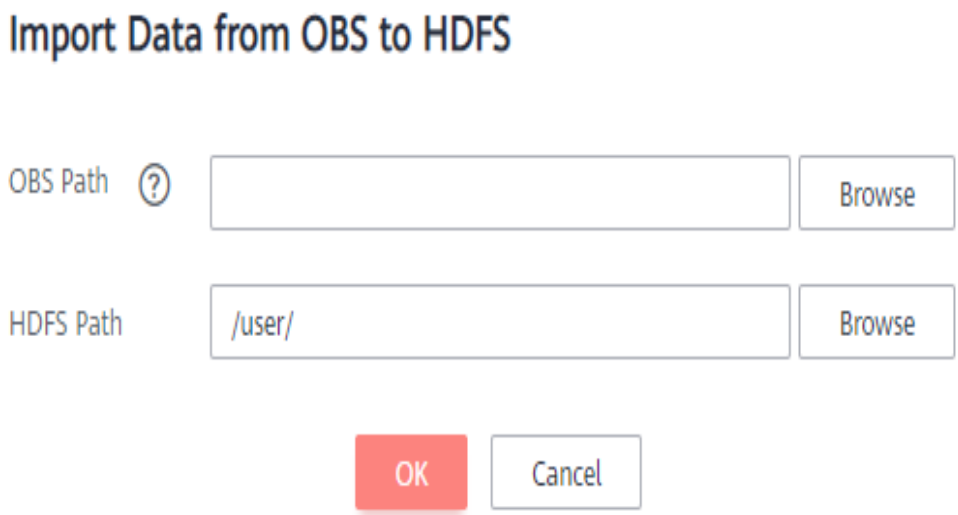
   - The folder name contains a maximum of 255 characters.
   - The folder name cannot be empty.
   - The folder name cannot contain the following special characters: /:*?"<>|\;&,'`!{}[]$%+
   - The value cannot start or end with a period (.).

 Copyright © Huawei Technologies Co., Ltd.

- The spaces at the beginning and end are ignored.

6. Click **Import Data** and configure the HDFS and OBS paths correctly. When configuring the OBS or HDFS path, click **Browse**, select a file directory, and click **Yes**.

   Figure 1-1 Importing data

   - OBS path
     - The path must start with **obs://**.
     - Files or programs encrypted by KMS cannot be imported.
     - An empty folder cannot be imported.
     - The directory and file name can contain letters, digits, hyphens (-), and underscores (_), but cannot contain special characters ;|&>,<'$*?\
     - The directory and file name cannot start or end with a space, but can contain spaces between them.
     - The OBS full path contains a maximum of 255 characters.
   - HDFS path
     - The path starts with **/user** by default.
     - The directory and file name can contain letters, digits, hyphens (-), and underscores (_), but cannot contain the following special characters: ;|&>,<'$*?\:
     - The directory and file name cannot start or end with a space, but can contain spaces between them.
     - The HDFS full path contains a maximum of 255 characters.

7. Click **OK**.

 Copyright © Huawei Technologies Co., Ltd.

You can view the file upload progress on the **File Operation Records** page. MRS processes the data import operation as a DistCp job. You can also check whether the DistCp job is successfully executed on the **Jobs** page.

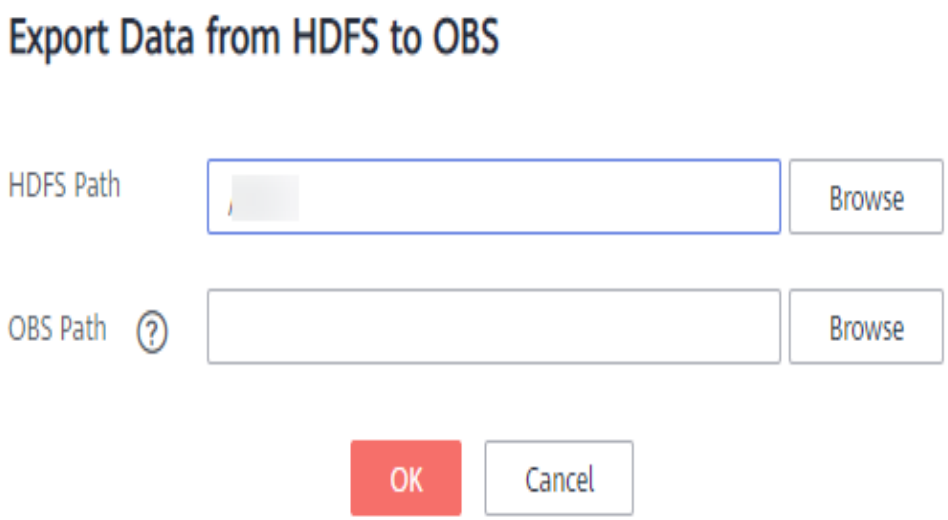
Exporting Data

After data analysis and computing is complete, you can store the data in the HDFS or export it to OBS.

You can perform the following steps to export files and directories:

1. Log in to the MRS console.
2. Choose **Clusters** > **Active Clusters**, and click the name of the target cluster to enter the cluster details page.
3. Click **Files** to go to the file management page.
4. Select **HDFS File List**.
5. Go to the data storage directory, for example, **bd_app1**.
6. Click **Export Data** and configure the OBS and HDFS paths. When configuring the OBS or HDFS path, click **Browse**, select a file directory, and click **Yes**.

   Figure 1-2 Exporting data



   - OBS path
     - The path must start with **obs://**.
     - The directory and file name can contain letters, digits, hyphens (-), and underscores (_), but cannot contain special characters ;|&>,<'$*?\
     - The directory and file name cannot start or end with a space, but can contain spaces between them.
     - The OBS full path contains a maximum of 255 characters.

 Copyright © Huawei Technologies Co., Ltd.

- HDFS path
  - The path starts with **/user** by default.
  - The directory and file name can contain letters, digits, hyphens (-), and underscores (_), but cannot contain the following special characters: ;|&>,<'$*?\:
  - The directory and file name cannot start or end with a space, but can contain spaces between them.
  - The HDFS full path contains a maximum of 255 characters.

NOTE

When a folder is exported to OBS, a label file named **folder name_$folder$** is added to the OBS path. Ensure that the exported folder is not empty. If the exported folder is empty, OBS cannot display the folder and only generates a file named **folder name_$folder$**.

7. Click **OK**.

   You can view the file upload progress on the **File Operation Records** page. MRS processes the data export operation as a DistCp job. You can also check whether the DistCp job is successfully executed on the **Jobs** page.

1.4 Creating a Job

You can submit programs developed by yourself to MRS to execute them, and obtain the results.

This section describes how to submit a job (take a MapReduce job as an example) on the MRS console. MapReduce jobs are used to submit JAR programs to quickly process massive amounts of data in parallel and create a distributed data processing and execution environment.

If the job and file management functions are not supported on the cluster details page, submit the jobs in the background.

Before creating a job, you need to upload local data to OBS for data computing and analyzing. MRS allows exporting data from OBS to HDFS for computing and analyzing. After the data analysis and computing are completed, you can store the data in HDFS or export them to OBS. HDFS and OBS can also store the compressed data in the format of **bz2** or **gz**.

NOTE

If the IAM username contains spaces (for example, **admin 01**), a job cannot be created.

Submitting a Job on the GUI

Step 1 Log in to the MRS console.

Step 2 Choose **Clusters** > **Active Clusters**, select a running cluster, and click its name to access the cluster details page.

Step 3 If Kerberos authentication is enabled for the cluster, perform the following steps. If Kerberos authentication is not enabled for the cluster, skip this step.

 Copyright © Huawei Technologies Co., Ltd.

In the **Basic Information** area on the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.

Step 4 Click the **Jobs** tab.

Step 5 Click **Create**. The **Create Job** dialog box is displayed.

Step 6 In **Type**, select **MapReduce**. Configure other job information.

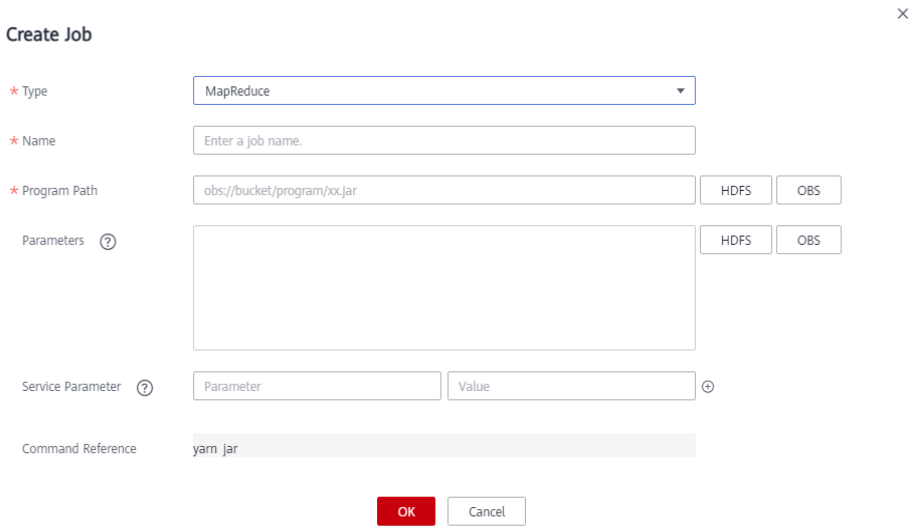


Table 1-1 Job parameters

Parameter	Description
Name	Job name. It contains 1 to 64 characters. Only letters, digits, hyphens (-), and underscores (_) are allowed. **NOTE** You are advised to set different names for different jobs.
Program Path	Path of the program package to be executed. The following requirements must be met: • Contains a maximum of 1,023 characters, excluding special characters such as **;|&><'$**. The parameter value cannot be empty or full of spaces. • The path of the program to be executed can be stored in HDFS or OBS. The path varies depending on the file system. – OBS: The path starts with **obs://**. Example: **obs://wordcount/program/*xxx*.jar** – HDFS: The path must start with **/user**. • For SparkScript and HiveScript, the path must end with **.sql**. For MapReduce, the path must end with **.jar**. For Flink and SparkSubmit, the path must end with **.jar** or **.py**. The **.sql**, **.jar**, and **.py** are case-insensitive.

 Copyright © Huawei Technologies Co., Ltd.

Parameter	Description
Parameters	(Optional) It is the key parameter for program execution. Separate multiple parameters with space. Configuration method: *Program class name Data input path Data output path* • Program class name: It is specified by a function in your program. MRS is responsible for transferring parameters only. • Data input path: Click **HDFS** or **OBS** to select a path or manually enter a correct path. • Data output path: Enter a directory that does not exist. The parameter contains a maximum of 150,000 characters. It cannot contain special characters ;|&><'$, but can be left blank. **CAUTION** If you enter a parameter with sensitive information (such as the login password), the parameter may be exposed in the job details display and log printing. Exercise caution when performing this operation.
Service Parameters	(Optional) Used to modify service configuration parameters for the job to be executed. The parameter modification applies only to the job to be executed. To add multiple parameters, click ⊕ on the right. To delete a parameter, click **Delete** on the right. **Table 1-2** describes the common parameters of a service.
Command Reference	Command submitted to the background for execution when a job is submitted.

Table 1-2 Service configuration parameters

Parameter	Description	Example Value
fs.obs.access.key	Key ID for accessing OBS.	-
fs.obs.secret.key	Key corresponding to the key ID for accessing OBS.	-

Step 7 Confirm job configuration information and click **OK**.

After the job is created, you can manage it.

----End

 Copyright © Huawei Technologies Co., Ltd.

1.5 Terminating a Cluster

You can terminate an MRS cluster that is no longer use after job execution is complete. The terminated or unsubscribed cluster is no longer billed.

Background

Typically after data is analyzed and stored, or when the cluster encounters an exception and cannot work, you can terminate a cluster. A cluster failed to be deployed will be automatically terminated.

Procedure

Step 1 Log in to the MRS management console.

Step 2 In the navigation pane on the left, choose **Clusters** > **Active Clusters**.

Step 3 In the cluster list, locate the row containing the cluster to be terminated, and click **Terminate** in the **Operation** column.

The cluster status changes from **Running** to **Terminating**, and finally to **Terminated**. You can view the terminated cluster in **Cluster History**. The terminated cluster is no longer billed.

----End

 Copyright © Huawei Technologies Co., Ltd.

2 Installing and Using the Cluster Client

This section describes how to quickly install and use clients of all services in an MRS 3.x cluster or later.

Clients can be installed on the nodes either in or outside the cluster. The following provides an example of how to install and use a client in a cluster.

NOTE

If Flume has been installed in the cluster, the Flume client must be installed independently. For details about how to install the Flume client, see **Installing the Flume Client**.

You can get started by reading the following topics:

1. **Downloading a Client**
2. **Installing a Client**
3. **Using a Client**

Downloading a Client

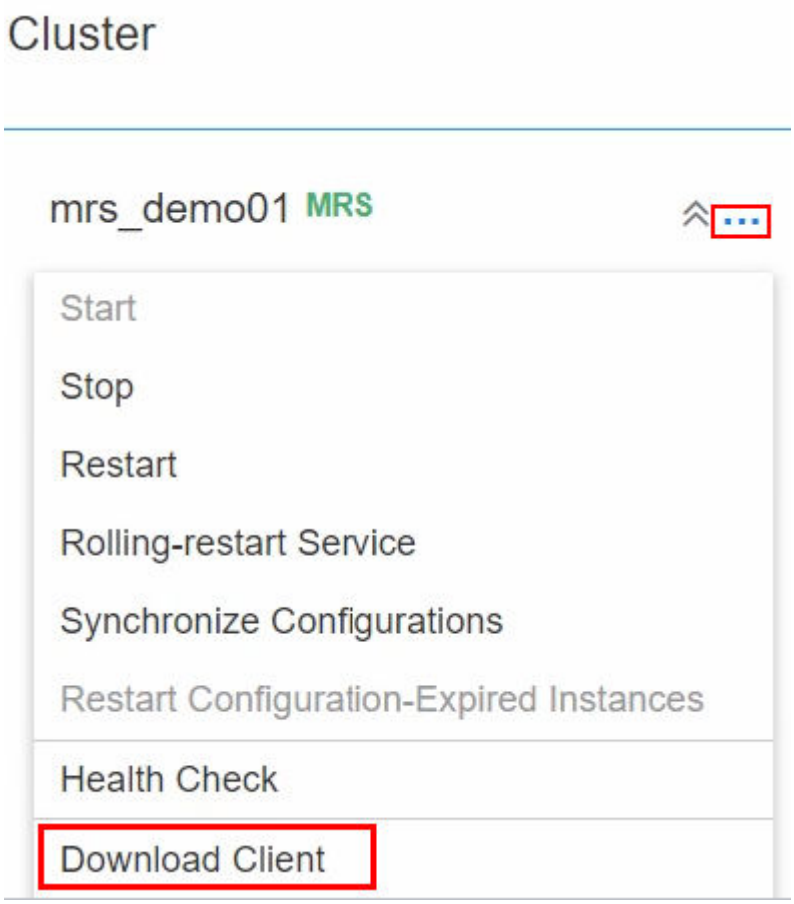
Step 1 Log in to FusionInsight Manager of the cluster by referring to **Accessing FusionInsight Manager (MRS 3.x or Later)**.

Step 2 Download the software package of the cluster client to the target node.

On the home page, click ··· next to the cluster name and click **Download Client** to download the cluster client.

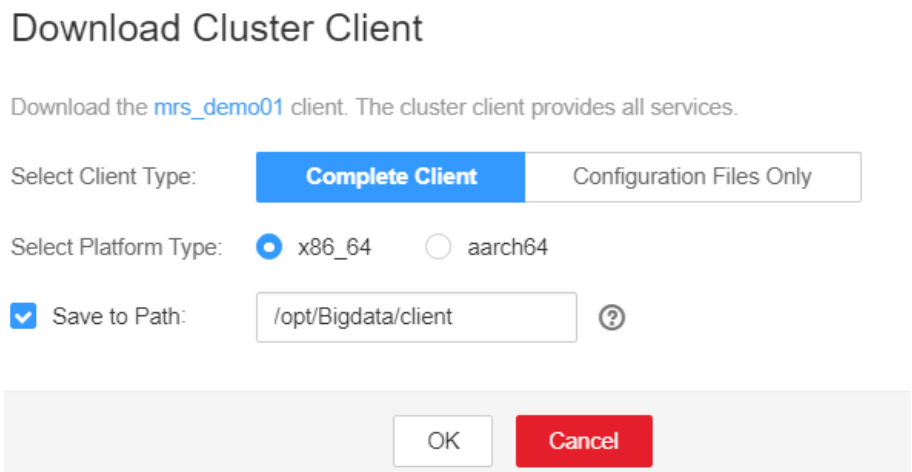
 Copyright © Huawei Technologies Co., Ltd.

Figure 2-1 Downloading a client



Step 3 On the **Download Cluster Client** page, enter the cluster client download information.

Figure 2-2 Downloading the cluster client



- Set **Select Client Type** to **Complete Client**.
- Set **Select Platform Type** to the architecture of the node to install the client. **x86_64** is used as an example.

 Copyright © Huawei Technologies Co., Ltd.

- Select **Save to Path** and enter the download path, for example, **/opt/Bigdata/client**. Ensure that user **omm** has the operation permission on the path.

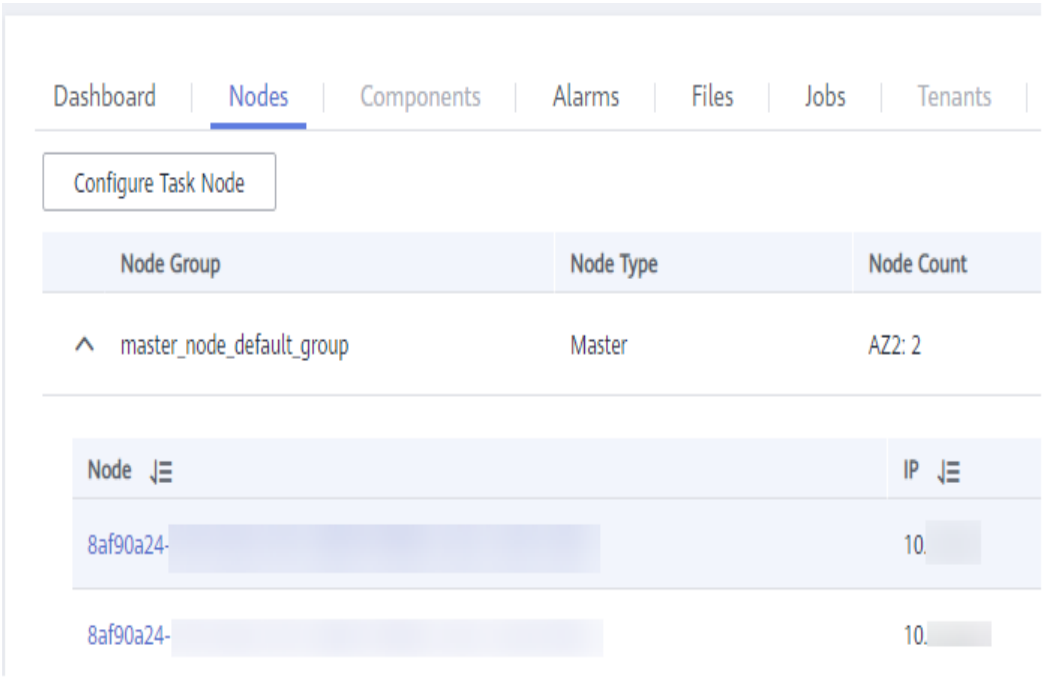
  NOTE

  The cluster supports two types of clients: **x86_64** and **aarch64**. The client type must match the architecture of the node for installing the client. Otherwise, client installation will fail.

Step 4 After the client software package is downloaded, log in to the active OMS node of the cluster as user **root**.

By default, the client software package is downloaded to the active OMS node of the cluster. You can view the node marked with ★ on the host page of FusionInsight Manager. If you need to install the client software package on another node in the cluster, run the following command to transfer the software package to the target node.

In the cluster list on the MRS console, click the cluster name. On the **Nodes** page, click the name of the target node. On the ECS details page, you can remotely log in to this node.



scp -p /opt/Bigdata/client/FusionInsight_Cluster_1_Services_Client.tar *IP address of the node where the client is to be installed:/opt/Bigdata/client*

----End

Installing a Client

Step 1 Log in to the node where the client software package is installed as the client user (for example, user **root**) and run the following commands to decompress the software package:

cd /opt/Bigdata/client

 Copyright © Huawei Technologies Co., Ltd.

tar -xvf FusionInsight_Cluster_1_Services_Client.tar

Step 2 Run the **sha256sum** command to verify the decompressed file.

sha256sum -c FusionInsight_Cluster_1_Services_ClientConfig.tar.sha256

```
FusionInsight_Cluster_1_Services_Client.tar: OK
```

Step 3 Decompress the obtained installation file.

tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar

Step 4 Go to the directory where the installation package is stored and install the client.

cd /opt/Bigdata/client/FusionInsight_Cluster_1_Services_ClientConfig

Run the following command to install the client to a specified directory (an absolute path), for example, **/opt/hadoopclient**.

./install.sh /opt/hadoopclient

```
...
The component client is installed successfully
```

NOTE

- If the **/opt/hadoopclient** directory has been used by existing service clients, you need to use another directory in this step when installing other service clients.
- You must delete the client installation directory when uninstalling a client.
- If you want to prevent other users from accessing this client, add parameter **-o** during the installation. That is, run the **./install.sh /opt/hadoopclient -o** command to install the client.
- If an HBase client is installed, it is recommended that the client installation directory contain only uppercase and lowercase letters, digits, and special characters (_-?.@+=) due to the limitation of the Ruby syntax used by HBase.
- If the NTP server is to be installed in **chrony** mode, ensure that the parameter **chrony** is added during the installation, that is, run the **./install.sh /opt/hadoopclient -o chrony** command to install the client.

----End

Using a Client

Step 1 Log in to the node where the client is installed as the client installation user, and run the following command to switch to the client directory:

cd /opt/hadoopclient

Step 2 Run the following command to load environment variables:

source bigdata_env

Step 3 If Kerberos authentication is enabled for the current cluster, run the following command to authenticate the user. If Kerberos authentication is disabled for the current cluster, authentication is not required.

kinit *MRS cluster user*

For example:

kinit admin

 Copyright © Huawei Technologies Co., Ltd.

Step 4 Run the client command of a component directly.

For example:

Run the following command to view files in the HDFS root directory:

hdfs dfs -ls /

```
Found 15 items
drwxrwx--x   - hive       hive              0 2021-10-26 16:30 /apps
drwxr-xr-x   - hdfs       hadoop            0 2021-10-18 20:54 /datasets
drwxr-xr-x   - hdfs       hadoop            0 2021-10-18 20:54 /datastore
drwxrwx---+  - flink      hadoop            0 2021-10-18 21:10 /flink
drwxr-x---   - flume      hadoop            0 2021-10-18 20:54 /flume
drwxrwx--x   - hbase      hadoop            0 2021-10-30 07:31 /hbase
...
```

----End

 Copyright © Huawei Technologies Co., Ltd.

3 Using Clusters with Kerberos Authentication Enabled

This section instructs you to use security clusters and run MapReduce, Spark, and Hive programs.

The Presto component of MRS 3.x does not support Kerberos authentication.

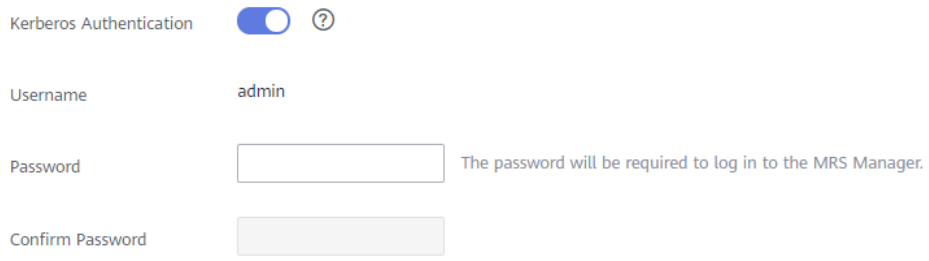
You can get started by reading the following topics:

1. **Creating a Security Cluster and Logging In to Manager**
2. **Creating a Role and a User**
3. **Running a MapReduce Program**
4. **Running a Spark Program**
5. **Running a Hive Program**

Creating a Security Cluster and Logging In to Manager

Step 1 Create a security cluster. For details, see **Custom Purchase of a Cluster**. Enable **Kerberos Authentication**, configure **Password**, and confirm the password. This password is used to log in to Manager. Keep it secure.

Figure 3-1 Setting security cluster parameters



Step 2 Log in to the MRS console.

Step 3 In the navigation pane on the left, choose **Active Clusters** and click the target cluster name on the right to access the cluster details page.

 Copyright © Huawei Technologies Co., Ltd.

Step 4 Click **Access Manager** on the right of **MRS Manager** to log in to Manager.

- If you have bound an EIP when creating the cluster, perform the following operations:

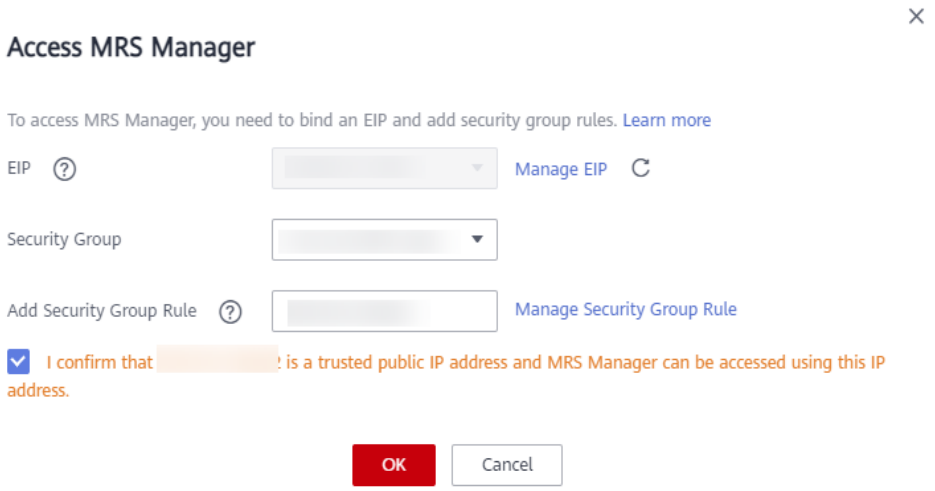
  a. Add a security group rule. By default, your public IP address used for accessing port 9022 is filled in the rule. If you want to view, modify, or delete a security group rule, click **Manage Security Group Rule**.

  NOTE

  - It is normal that the automatically generated public IP address is different from your local IP address and no action is required.
  - If port 9022 is a Knox port, you need to enable the permission to access port 9022 of Knox for accessing Manager.

  b. Select **I confirm that xx.xx.xx.xx is a trusted public IP address and MRS Manager can be accessed using this IP address.**

  Figure 3-2 Accessing Manager



- If you have not bound an EIP when creating the cluster, perform the following operations:

  a. Select an EIP from the drop-down list or click **Manage EIP** to buy one.

  b. Add a security group rule. By default, your public IP address used for accessing port 9022 is filled in the rule. If you want to view, modify, or delete a security group rule, click **Manage Security Group Rule**.

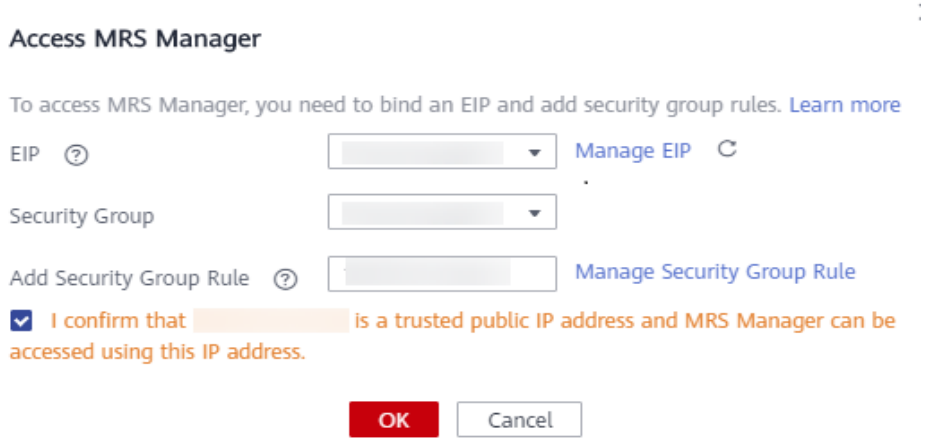
  NOTE

  - It is normal that the automatically generated public IP address is different from the local IP address and no action is required.
  - If port 9022 is a Knox port, you need to enable the permission of port 9022 to access Knox for accessing MRS Manager.

  c. Select **I confirm that xx.xx.xx.xx is a trusted public IP address and MRS Manager can be accessed using this IP address.**

 Copyright © Huawei Technologies Co., Ltd.

Figure 3-3 Accessing Manager



Step 5 Click **OK**. The Manager login page is displayed. To assign other users the permission to access Manager, add the IP addresses as trusted ones by referring to **Accessing Manager**.

Step 6 Enter the default username **admin** and the password you set when creating the cluster, and click **Log In**.

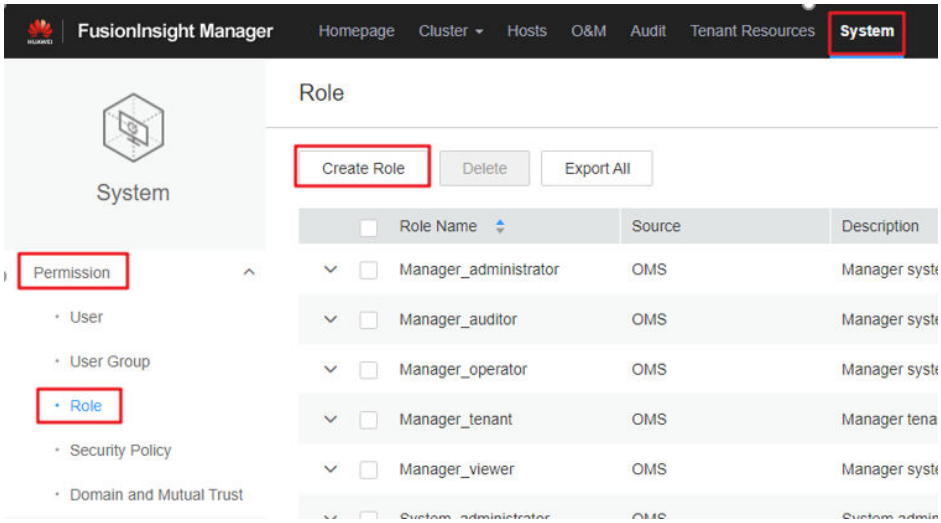
----End

Creating a Role and a User

For clusters with Kerberos authentication enabled, perform the following steps to create a user and assign permissions to the user to run programs.

Step 1 On Manager, choose **System** > **Permission** > **Role**.

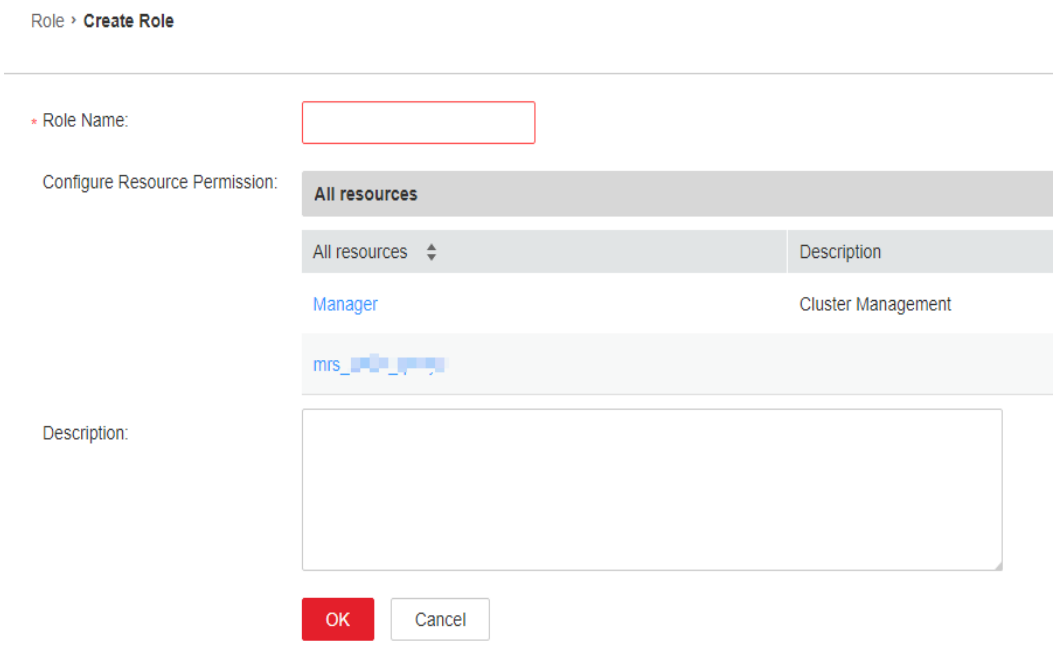
Figure 3-4 Role



Step 2 Click **Create Role**. For details, see **Creating a Role**.

 Copyright © Huawei Technologies Co., Ltd.

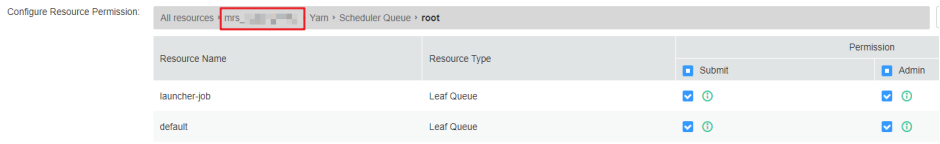
Figure 3-5 Creating a role



Specify the following information:

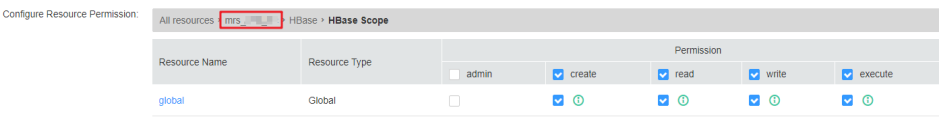
- Enter a role name, for example, **mrrole**.
- In **Configure Resource Permission**, select the cluster to be operated, choose **Yarn** > **Scheduler Queue** > **root**, and select **Submit** and **Admin** in the **Permission** column. After you finish configuration, do not click **OK** but click the name of the target cluster shown in the following figure and then configure other permissions.

  Figure 3-6 Configuring resource permissions for Yarn

- Choose **HBase** > **HBase Scope**. Locate the row that contains **global**, and select **create**, **read**, **write**, and **execute** in the **Permission** column. After you finish configuration, do not click **OK** but click the name of the target cluster shown in the following figure and then configure other permissions.

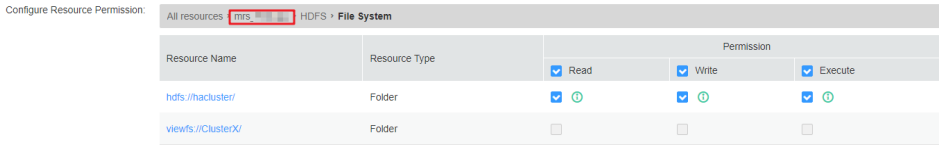
  Figure 3-7 Configuring resource permissions for HBase

- Choose **HDFS** > **File System** > **hdfs://hacluster/** and select **Read**, **Write**, and **Execute** in the **Permission** column. After you finish configuration, do not click **OK** but click the name of the target cluster shown in the following figure and then configure other permissions.

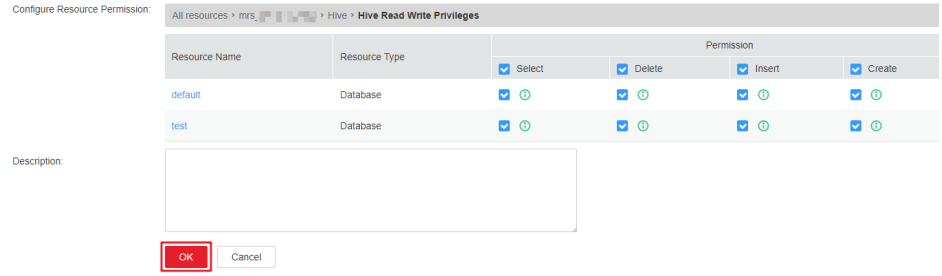
 Copyright © Huawei Technologies Co., Ltd.

Figure 3-8 Configuring resource permissions for HDFS



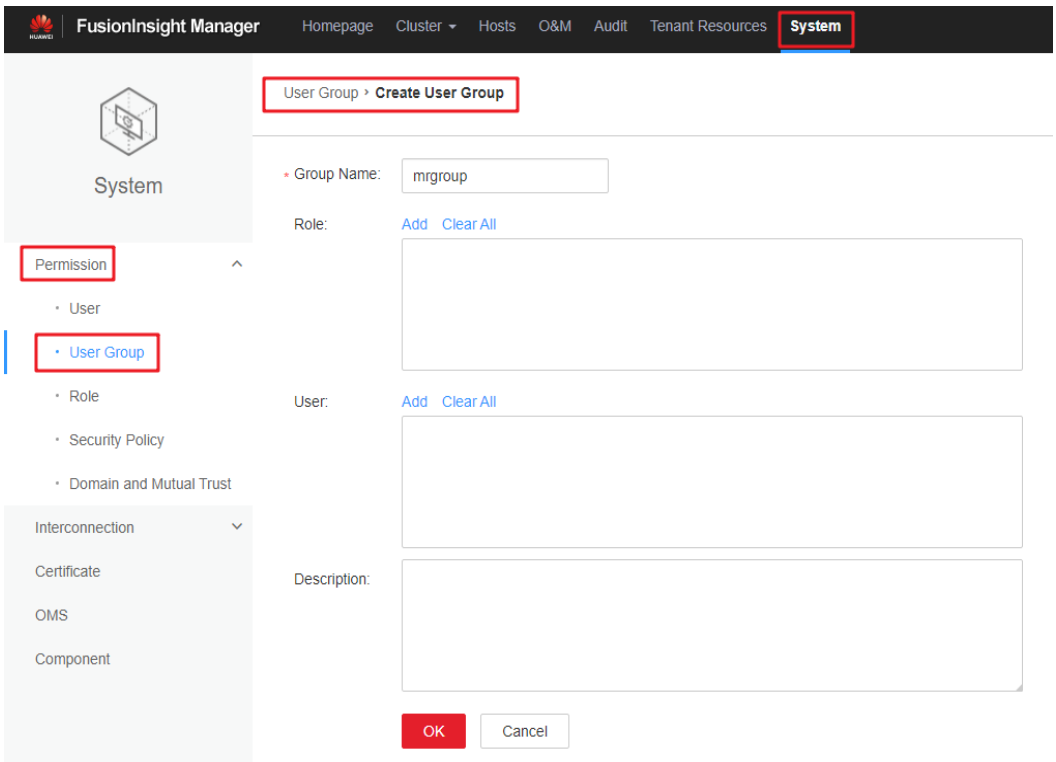
- Choose **Hive** > **Hive Read Write Privileges**, select **Select**, **Delete**, **Insert**, and **Create** in the **Permission** column, and click **OK**.

Figure 3-9 Configuring resource permissions for Hive



Step 3 Choose **System**. In the navigation pane on the left, choose **Permission** > **User Group** > **Create User Group** to create a user group for the sample project, for example, **mrgroup**. For details, see **Creating a User Group**.

Figure 3-10 Creating a user group



Step 4 Choose **System**. In the navigation pane on the left, choose **Permission** > **User** > **Create** to create a user for the sample project. For details, see **Creating a User**.

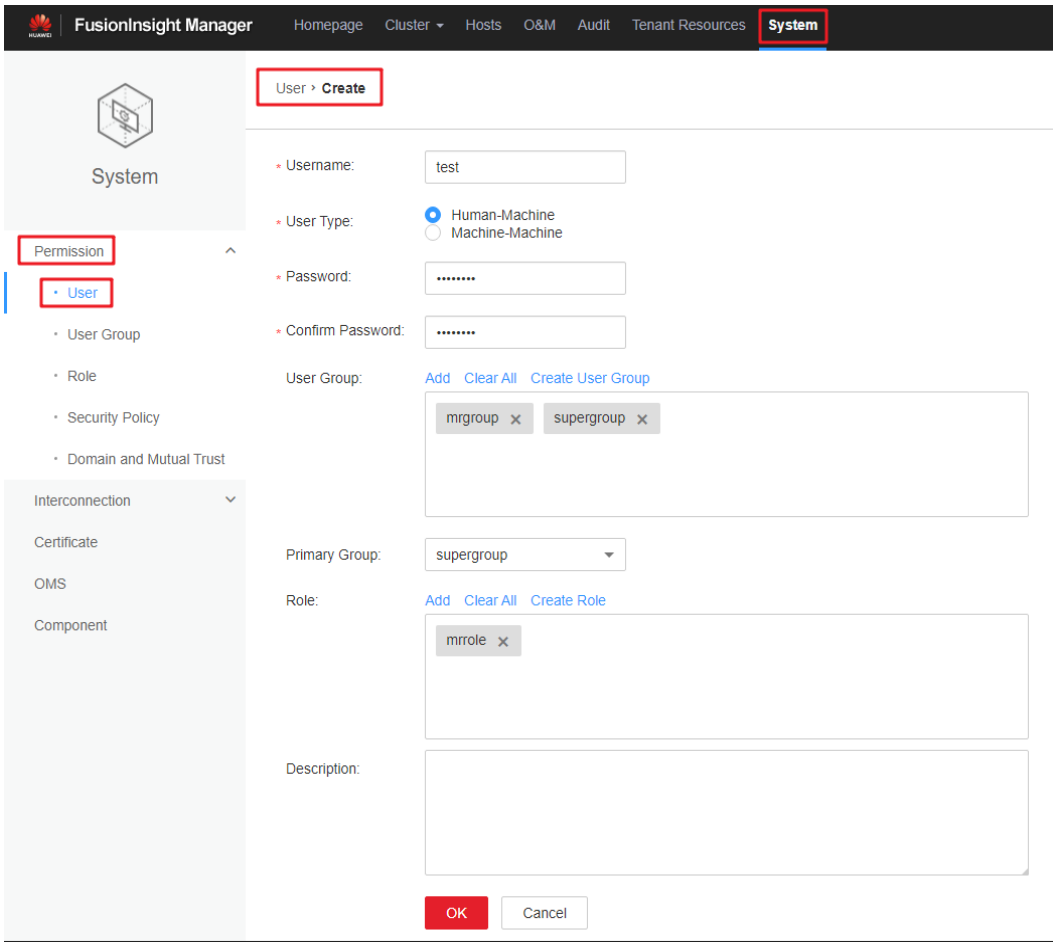
- Enter a username, for example, **test**. If you want to run a Hive program, enter **hiveuser** in **Username**.

 Copyright © Huawei Technologies Co., Ltd.

- Set **User Type** to **Human-Machine**.
- Enter a password. This password will be used when you run the program.
- In **User Group**, add **mrgroup** and **supergroup**.
- Set **Primary Group** to **supergroup** and bind the **mrrole** role to obtain the permission.

  Click **OK**.

Figure 3-11 Creating a user



Step 5 Choose **System**. In the navigation pane on the left, choose **Permission** > **User**, locate the row where user **test** locates, and select **Download Authentication Credential** from the **More** drop-down list. Save the downloaded package and decompress it to obtain the **keytab** and **krb5.conf** files.

Figure 3-12 Downloading the authentication credential

----End

 Copyright © Huawei Technologies Co., Ltd.

Running a MapReduce Program

This section describes how to run a MapReduce program in security cluster mode.

Prerequisites

You have compiled the program and prepared data files, for example, **mapreduce-examples-1.0.jar**, **input_data1.txt**, and **input_data2.txt**. For details about MapReduce program development and data preparations, see **MapReduce Introduction**.

Procedure

Step 1 Use a remote login software (for example, MobaXterm) to log in to the master node of the security cluster using SSH (using the EIP).

Step 2 After the login is successful, run the following commands to create the **test** folder in the **/opt/Bigdata/client** directory and create the **conf** folder in the **test** directory:

```
cd /opt/Bigdata/client
mkdir test
cd test
mkdir conf
```

Step 3 Use an upload tool (for example, WinSCP) to copy **mapreduce-examples-1.0.jar**, **input_data1.txt**, and **input_data2.txt** to the **test** directory, and copy the **keytab** and **krb5.conf** files obtained in **Step 5** in **Creating Roles and Users** to the **conf** directory.

Step 4 Run the following commands to configure environment variables and authenticate the created user, for example, **test**:

```
cd /opt/Bigdata/client
source bigdata_env
export YARN_USER_CLASSPATH=/opt/Bigdata/client/test/conf/
kinit test
```

Enter the password as prompted. If no error message is displayed (you need to change the password as prompted upon the first login), Kerberos authentication is complete.

Step 5 Run the following commands to import data to the HDFS:

```
cd test
hdfs dfs -mkdir /tmp/input
hdfs dfs -put input_data* /tmp/input
```

Step 6 Run the following commands to run the program:

```
yarn jar mapreduce-examples-1.0.jar com.huawei.bigdata.mapreduce.examples.FemaleInfoCollector /tmp/input /tmp/mapreduce_output
```

In the preceding commands:

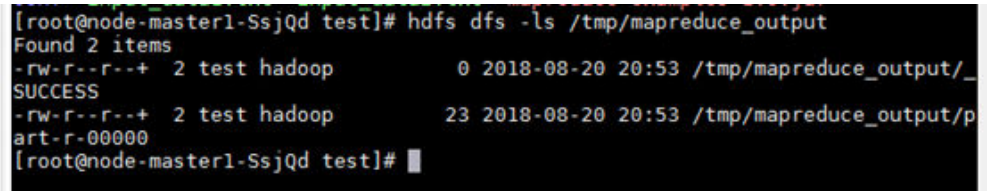
/tmp/input indicates the input path in the HDFS.

/tmp/mapreduce_output indicates the output path in the HDFS. This directory must not exist. Otherwise, an error will be reported.

Step 7 After the program is executed successfully, run the **hdfs dfs -ls /tmp/mapreduce_output** command. The following command output is displayed.

 Copyright © Huawei Technologies Co., Ltd.

Figure 3-13 Program running result



----End

Running a Spark Program

This section describes how to run a Spark program in security cluster mode.

Prerequisites

You have compiled the program and prepared data files, for example, **FemaleInfoCollection.jar**, **input_data1.txt**, and **input_data2.txt**. For details about Spark program development and data preparations, see **Spark Application Development Overview**.

Procedure

Step 1 Use a remote login software (for example, MobaXterm) to log in to the master node of the security cluster using SSH (using the EIP).

Step 2 After the login is successful, run the following commands to create the **test** folder in the **/opt/Bigdata/client** directory and create the **conf** folder in the **test** directory:

```
cd /opt/Bigdata/client
mkdir test
cd test
mkdir conf
```

Step 3 Use an upload tool (for example, WinSCP) to copy **FemaleInfoCollection.jar**, **input_data1.txt**, and **input_data2.txt** to the **test** directory, and copy the **keytab** and **krb5.conf** files obtained in **Step 5** in section **Creating Roles and Users** to the **conf** directory.

Step 4 Run the following commands to configure environment variables and authenticate the created user, for example, **test**:

```
cd /opt/Bigdata/client
source bigdata_env
export YARN_USER_CLASSPATH=/opt/Bigdata/client/test/conf/
kinit test
```

Enter the password as prompted. If no error message is displayed, Kerberos authentication is complete.

Step 5 Run the following commands to import data to the HDFS:

```
cd test
hdfs dfs -mkdir /tmp/input
hdfs dfs -put input_data* /tmp/input
```

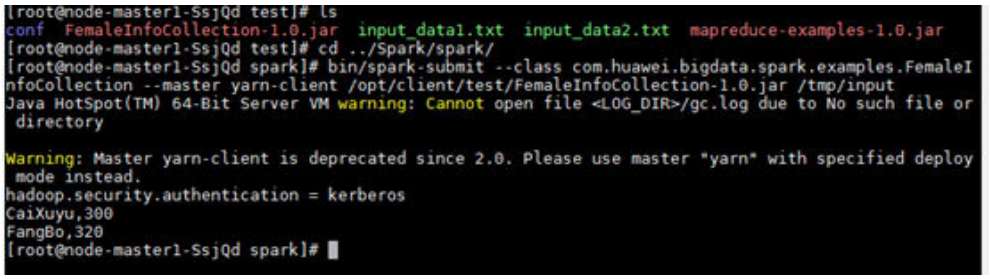
Step 6 Run the following commands to run the program:

```
cd /opt/Bigdata/client/Spark/spark
bin/spark-submit --class com.huawei.bigdata.spark.examples.FemaleInfoCollection --master yarn-client /opt/Bigdata/client/test/FemaleInfoCollection-1.0.jar /tmp/input
```

 Copyright © Huawei Technologies Co., Ltd.

Step 7 After the program is run successfully, the following information is displayed.

Figure 3-14 Program running result



----End

Running a Hive Program

This section describes how to run a Hive program in security cluster mode.

Prerequisites

You have compiled the program and prepared data files, for example, **hive-examples-1.0.jar**, **input_data1.txt**, and **input_data2.txt**. For details about Hive program development and data preparations, see **Hive Application Development Overview**.

Procedure

Step 1 Use a remote login software (for example, MobaXterm) to log in to the master node of the security cluster using SSH (using the EIP).

Step 2 After the login is successful, run the following commands to create the **test** folder in the **/opt/Bigdata/client** directory and create the **conf** folder in the **test** directory:

```
cd /opt/Bigdata/client
mkdir test
cd test
mkdir conf
```

Step 3 Use an upload tool (for example, WinSCP) to copy **FemaleInfoCollection.jar**, **input_data1.txt**, and **input_data2.txt** to the **test** directory, and copy the **keytab** and **krb5.conf** files obtained in **Step 5** in section **Creating Roles and Users** to the **conf** directory.

Step 4 Run the following commands to configure environment variables and authenticate the created user, for example, **test**:

```
cd /opt/Bigdata/client
source bigdata_env
export YARN_USER_CLASSPATH=/opt/Bigdata/client/test/conf/
kinit test
```

Enter the password as prompted. If no error message is displayed, Kerberos authentication is complete.

Step 5 Run the following command to run the program:

```
chmod +x /opt/hive_examples -R  cd /opt/hive_examples  java -cp .:hive-examples-1.0.jar:/opt/
hive_examples/conf:/opt/Bigdata/client/Hive/Beeline/lib/*:/opt/Bigdata/client/HDFS/hadoop/lib/*
com.huawei.bigdata.hive.example.ExampleMain
```

 Copyright © Huawei Technologies Co., Ltd.

Step 6 After the program is run successfully, the following information is displayed.

Figure 3-15 Program running result

----End

 Copyright © Huawei Technologies Co., Ltd.

4 Using Hadoop from Scratch

- MapReduce Service (MRS) provides Hadoop-based high-performance big data components, such as Spark, HBase, Kafka, and Storm.
- This section describes how to use Hadoop to submit wordcount jobs through the GUI and cluster nodes. A wordcount job is the most classic Hadoop job that counts words in massive amounts of text.
- Purchase a cluster; prepare the Hadoop sample program and data files; upload data to OBS; create a job; and view job execution results.

  You can get started by reading the following steps:

  a. **Purchase an MRS cluster.**
  b. **Configure software.**
  c. **Configure hardware.**
  d. **Set advanced options.**
  e. **Prepare the Hadoop sample program and data files.**
  f. **Upload data to OBS.**
  g. **Submit a job on the GUI.**
  h. **Submit a job through a cluster node.**
  i. **Query job execution results.**

Procedure

Step 1 Purchase an MRS cluster.

1. Log in to the Huawei Cloud console.
2. Choose **Service List** > **Analytics** > **MapReduce Service**.
3. On the **Active Clusters** page that is displayed, click **Buy Cluster**.
4. Click the **Custom Config** tab.

Step 2 Configure software.

1. **Region**: Select a region as required.
2. **Cluster Name**: Enter **mrs_demo** or specify a name according to naming rules.
3. **Cluster Version**: Select **MRS 3.1.0**.
4. **Cluster Type**: Select **Analysis Cluster**.

 Copyright © Huawei Technologies Co., Ltd.

5. Select all analysis cluster components.
6. Click **Next**.

Step 3 Configure hardware.

1. **Billing Mode**: Select **Pay-per-use**.
2. **AZ**: Select **AZ2**.
3. **VPC** and **Subnet**: Retain their default values or click **View VPC** and **View Subnet** to create ones.
4. **Security Group**: Use the default value **Auto create**.
5. **EIP**: **Bind later** is selected by default.
6. **Enterprise Project**: Select **default**.
7. **Cluster Node**: Retain the default values. Do not add task nodes.
8. Click **Next**.

Step 4 Set advanced options.

1. **Tag**: Retain the default value.
2. **Agency**, **Alarm**, **Rule Name**, and **Topic Name**: Retain the default values.
3. **Kerberos Authentication**: **Disabled**
4. **Username**: **admin** is used by default.
5. **Password** and **Confirm Password**: Set them to the password of the FusionInsight Manager administrator.
6. **Login Mode**: Select **Password**. Enter a password and confirm the password for user **root**.
7. **Secure Communications**: Select **Enable**.
8. **Service Agreement**: Select **I have read and agree to the Huawei MRS Service Agreement**.
9. Click **Buy Now**. The page is displayed showing that the task has been submitted.
10. Click **Back to Cluster List**. You can view the status of the cluster on the **Active Clusters** page. Wait for the cluster creation to complete. The initial status of the cluster is **Starting**. After the cluster has been created, the cluster status becomes **Running**.

Step 5 Prepare the Hadoop sample program and data files.

1. Prepare the wordcount program.

   Download the **Hadoop sample program** (including wordcount). **hadoop-3.1.4.tar.gz** is used as an example. Use the actual program version provided in the link. For example, choose **hadoop-3.1.4**. On the page that is displayed, click **hadoop-3.1.4.tar.gz** to download it. Then, decompress it to obtain **hadoop-3.1.4\share\hadoop\mapreduce** (the Hadoop sample program) from **hadoop-mapreduce-examples-3.1.4.jar**.

2. Prepare data files.

   There is no requirement on the format of data files. Prepare two **.txt** files. In this example, files **wordcount1.txt** and **wordcount2.txt** are used.

Step 6 Upload data to OBS.

 Copyright © Huawei Technologies Co., Ltd.

1. Log in to the OBS console and choose **Parallel File Systems**. On the **Parallel File Systems** page, click **Create Parallel File System**. On the **Create Parallel File System** page that is displayed, configure parameters to create a file system named **mrs-word01**.
2. Click the name of the **mrs-word01** file system. In the navigation pane on the left, choose **Files**. On the page that is displayed, click **Create Folder** to create the **program** and **input** folders.
3. Go to the **program** folder and upload the Hadoop sample program downloaded in **5**.
4. Go to the **input** folder and upload the **wordcount1.txt** and **wordcount2.txt** data files prepared in **5**.
5. To submit a job on the GUI, go to **7**.

   To submit a job through a cluster node, go to **8**.

Step 7 Submit a job on the GUI.

1. In the navigation pane of the MRS console, choose **Clusters** > **Active Clusters**. On the **Active Clusters** page, click the **mrs_demo** cluster.
2. On the cluster information page, click the **Jobs** tab then **Create** to create a job. To submit a job through a cluster node, go to **8**.
3. **Type**: **MapReduce**
4. **Job Name**: Enter **wordcount**.
5. **Program Path**: Click **OBS** and select the Hadoop sample program uploaded in **6**.
6. **Parameters**: Enter **wordcount obs://mrs-word01/input/ obs://mrs-word01/output/**. **output** indicates the output path. Enter a directory that does not exist.
7. **Service Parameters**: Leave it blank.
8. Click **OK** to submit the job. After a job is submitted, it is in the **Accepted** state by default. You do not need to manually execute the job.
9. Go to the **Jobs** tab page, view the job status and logs, and go to **9** to view the job execution result.

Step 8 Submit a job through a cluster node.

1. Log in to the MRS console and click the cluster named **mrs_demo** to go to its details page.
2. Click the **Nodes** tab. On this tab page, click the name of a master node to go to the ECS management console.
3. Click **Remote Login** in the upper right corner of the page.
4. Enter the username and password of the master node as prompted. The username is **root** and the password is the one configured during cluster creation.
5. Run the **source /opt/Bigdata/client/bigdata_env** command to configure environment variables.
6. If Kerberos authentication has been enabled, run the **kinit** *MRS cluster user* command, for example, **kinit admin**, to authenticate the current cluster user. Skip this step if Kerberos authentication is not enabled.
7. Run the following command to copy the sample program in the OBS bucket to the master node in the cluster:

 Copyright © Huawei Technologies Co., Ltd.

hadoop fs -Dfs.obs.access.key=AK -Dfs.obs.secret.key=SK -copyToLocal source_path.jar target_path.jar Example: **hadoop fs -Dfs.obs.access.key=XXXX -Dfs.obs.secret.key=XXXX -copyToLocal "obs://mrs-word01/program/hadoop-mapreduce-examples-XXX.jar" "/home/omm/hadoop-mapreduce-examples-XXX.jar"** To obtain the AK/SK pair for logging in to the OBS console, hover your cursor over the username in the upper right corner of the management console, and choose **My Credentials** > **Access Keys**, or click **Create Access Key** to create one.

8. Run the following command to submit a wordcount job. To read data from or write data to OBS, add AK/SK parameters. **source /opt/Bigdata/client/bigdata_env;hadoop jar execute_jar wordcount input_path output_path** Example: **source /opt/Bigdata/client/bigdata_env;hadoop jar /home/omm/hadoop-mapreduce-examples-XXX.jar wordcount -Dfs.obs.access.key=XXXX -Dfs.obs.secret.key=XXXX "obs://mrs-word01/input/*" "obs://mrs-word01/output/"** In this command, **input_path** indicates a path for storing job input files on OBS. **output_path** indicates a path for storing job output files on OBS and needs to be set to a directory that does not exist

Step 9 Query job execution results.

1. Log in to the OBS console and click the name of the **mrs-word01** bucket.
2. On the page that is displayed, choose **Objects** in the navigation pane on the left. Go to the output path in the **mrs-word01** bucket specified during job submission, and view the job output file. You need to download the file to the local host and open it in a **.txt** format.

----End

 Copyright © Huawei Technologies Co., Ltd.

5 Using Kafka from Scratch

MapReduce Service (MRS) provides Hadoop-based high-performance big data components, such as Spark, HBase, Kafka, and Storm.

This section uses a cluster with Kerberos authentication disabled as an example to describe how to generate and consume messages in a Kafka topic.

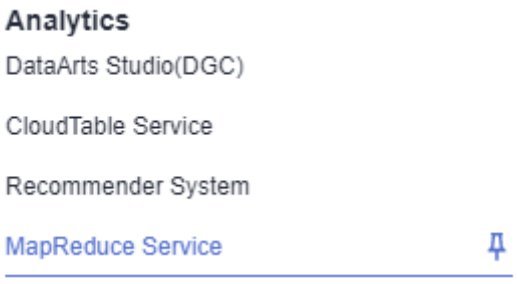
You can get started by reading the following steps:

1. **Purchasing a Cluster**
2. **Installing the Kafka Client**
3. **Logging In to a Master Node Using VNC**
4. **Creating a Topic Using the Kafka Client**
5. **Managing Messages in Kafka Topics**

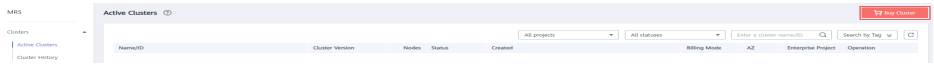
Purchasing a Cluster

Step 1 Purchase an MRS cluster.

1. Log in to the Huawei Cloud management console.
2. Choose **Service List** > **Analytics** > **MapReduce Service**.



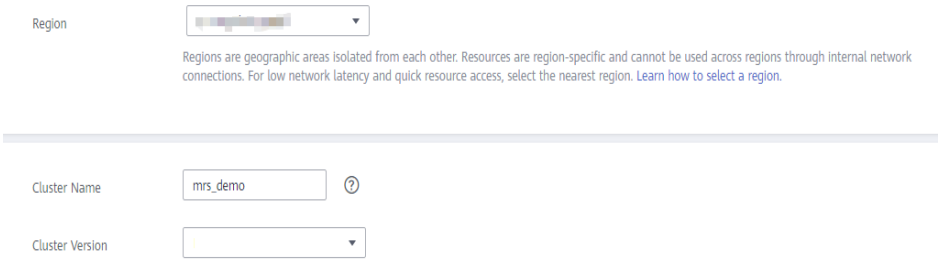
3. Choose **Clusters** > **Active Clusters**. On the **Active Clusters** page that is displayed, click **Buy Cluster**. On the displayed page, click the **Custom Config** tab.



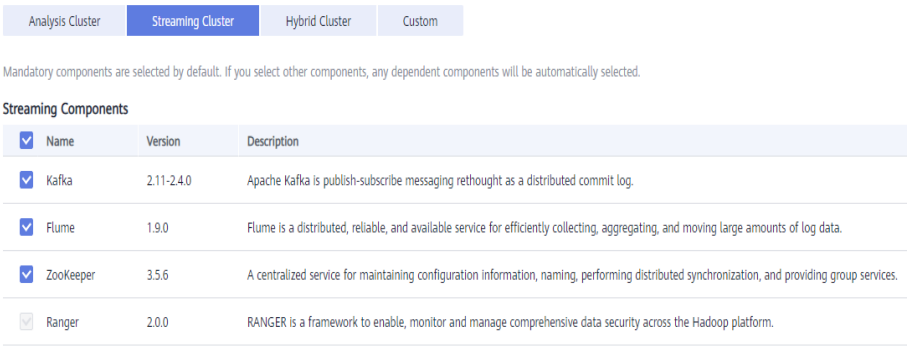
Step 2 Configure the following parameters and click **Next**.

 Copyright © Huawei Technologies Co., Ltd.

- **Region**: Select a region as required.
- **Cluster Name**: Enter **mrs_demo** or specify a name according to naming rules.
- **Cluster Version**: Select **MRS 3.1.0**.



- **Cluster Type**: Select **Streaming Cluster**, select all streaming cluster components, and click **Next**.



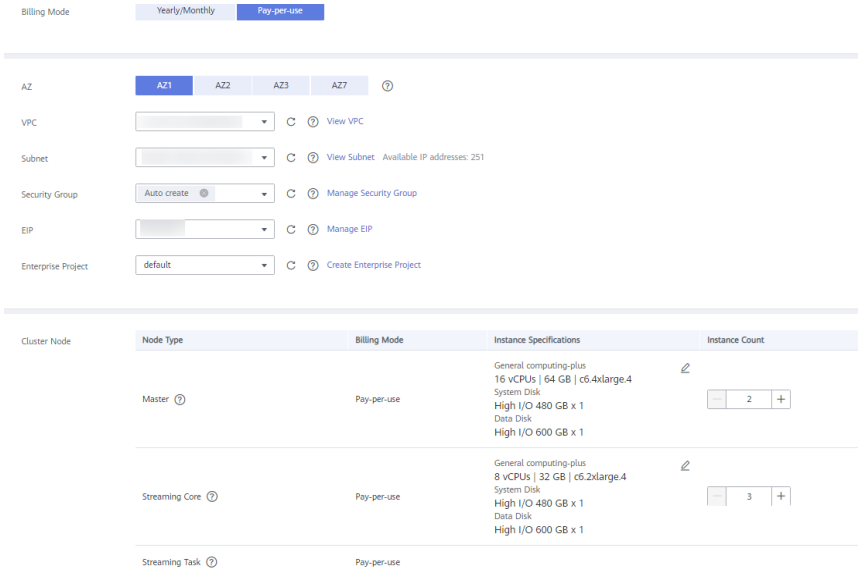
Step 3 On the **Configure Hardware** page, configure the parameters by referring to **Table 5-1**, and click **Next**.

Table 5-1 MRS hardware configuration parameters

Parameter	Example Value
Billing Mode	Pay-per-use
AZ	AZ2
VPC	Retain the default value. You can also click **View VPC** to create one.
EIP	You can select an existing EIP from the drop-down list. If no EIPs are available, click **Manage EIP** to access the **EIPs** page to create one.
Enterprise Project	default
Cluster Node	Retain the default values. Do not add task nodes.

 Copyright © Huawei Technologies Co., Ltd.

Figure 5-1 Configure Hardware



Step 4 Configure advanced options.

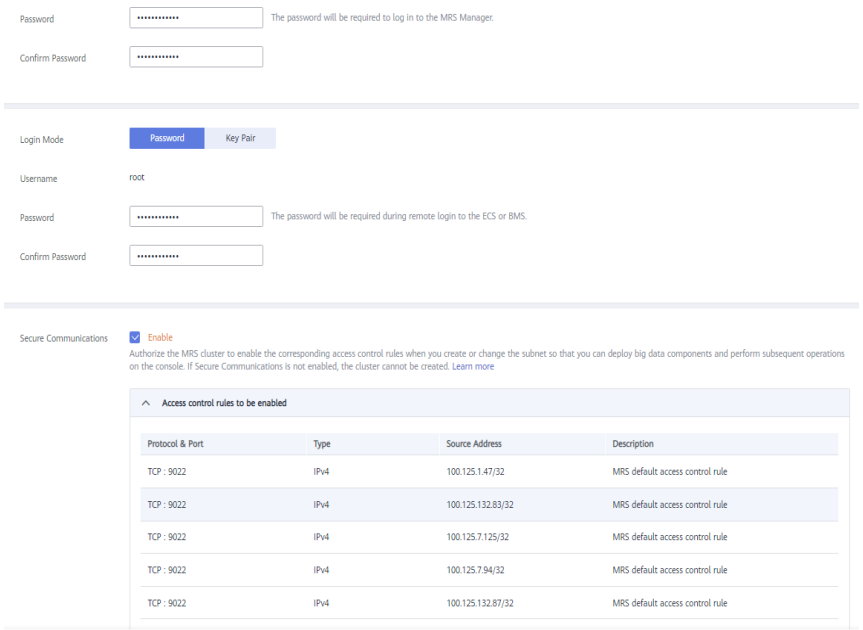
1. On the **Set Advanced Options** page, configure the parameters by referring to **Table 5-2** and retain the default values for other parameters.

Table 5-2 MRS cluster advanced parameters

Parameter	Example Value
Kerberos Authentication	Toggle the slider off.
Username	admin
Password	Configure the password for logging in to the cluster management page, for example, **Test!@12345**.
Confirm Password	Enter the password again.
Login Mode	Password
Username	root
Password	Configure a password for remotely logging in to ECSs or BMSs, for example, **Test!@12345**.
Confirm Password	Enter the password again.
Secure Communications	Select **Enable**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 5-2 Set Advanced Options



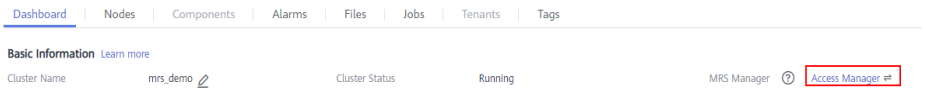
2. Click **Buy Now**. The page is displayed showing that the task has been submitted.
3. Click **Back to Cluster List**. You can view the status of the cluster on the **Active Clusters** page.
4. Wait for the cluster creation to complete. The initial status of the cluster is **Starting**. After the cluster has been created, the cluster status becomes **Running**.

----End

Installing the Kafka Client

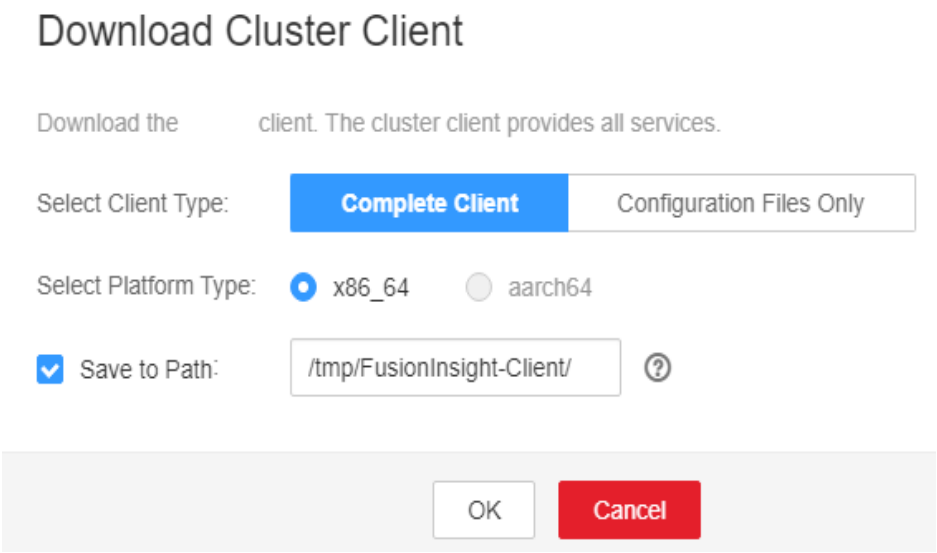
Step 1 Choose **Clusters** > **Active Clusters**. On the **Active Clusters** page, click the cluster named **mrs_demo** to go to its details page.

Step 2 Click **Access Manager** next to **MRS Manager**. On the page that is displayed, configure the EIP information and click **OK**. Enter the username and password to access FusionInsight Manager.



Step 3 Choose **Cluster** > **Services** > **HBase**. On the page displayed, choose **More** > **Download Client**. In the **Download Cluster Client** dialog box, select **Complete Client** for **Select Client Type**, select a platform type, select **Save to Path**, and click **OK**. The Kafka client software package, for example, **FusionInsight_Cluster_1_Kafka_Client.tar**, is downloaded.

 Copyright © Huawei Technologies Co., Ltd.



Step 4 Log in to the active node as user **root**.

Step 5 Go to the directory where the software package is stored and run the following commands to decompress and verify the software package, and decompress the obtained installation file:

cd /tmp/FusionInsight-Client

tar -xvf FusionInsight_Cluster_1_Kafka_Client.tar

sha256sum -c FusionInsight_Cluster_1_Kafka_ClientConfig.tar.sha256

tar -xvf FusionInsight_Cluster_1_Kafka_ClientConfig.tar

Step 6 Go to the directory where the installation package is stored, and run the following command to install the client to a specified directory (absolute path), for example, **/opt/hadoopclient**:

cd /tmp/FusionInsight-Client/FusionInsight_Cluster_1_Kafka_ClientConfig

Run the **./install.sh /opt/hadoopclient** command and wait until the client installation is complete.

Step 7 Check whether the client is installed.

cd /opt/hadoopclient

source bigdata_env

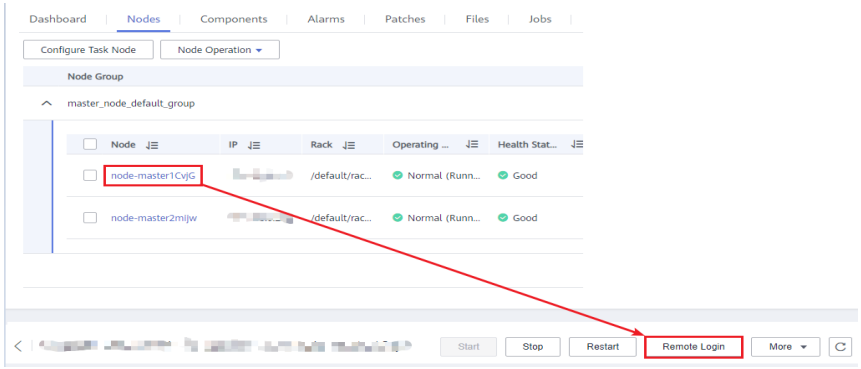
Run the **klist** command to query and confirm authentication details. If the command is executed, the Kafka client is installed.

----End

Logging In to a Master Node Using VNC

Step 1 Choose **Clusters** > **Active Clusters**. On the **Active Clusters** page that is displayed, click the cluster named **mrs_demo**. On the cluster details page that is displayed, click the **Nodes** tab. On this tab page, locate the node whose type is **Master1** and click the node name to go to the ECS details page.

 Copyright © Huawei Technologies Co., Ltd.



Step 2 Click **Remote Login** in the upper right corner of the page to remotely log in to the master node. Log in using the username **root** and the password configured during cluster purchase.

----End

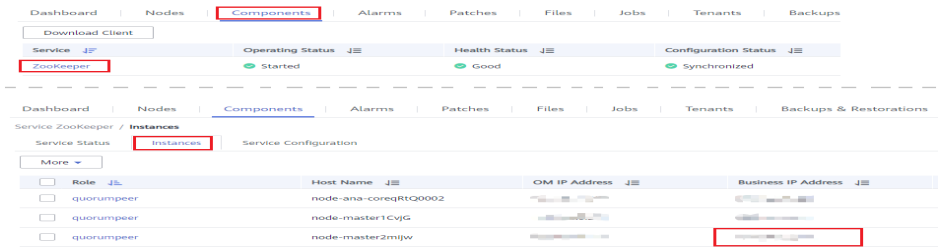
Creating a Topic Using the Kafka Client

Step 1 Configure environment variables. For example, if the Kafka client installation directory is **/opt/hadoopclient**, run the following command:

source /opt/hadoopclient/bigdata_env

Step 2 Choose **Clusters** > **Active Clusters**. On the **Active Clusters** page, click the cluster named **mrs_demo** to go to the **Dashboard** tab page. On this page, click **Synchronize** next to **IAM User Sync**.

Step 3 After the synchronization is complete, click the **Components** tab. On this tab page, select **ZooKeeper**. On the page that is displayed, click the **Instances** tab. Record the IP address of any ZooKeeper instance, for example, **192.168.7.35**.



Step 4 Run the following command to create a Kafka topic:

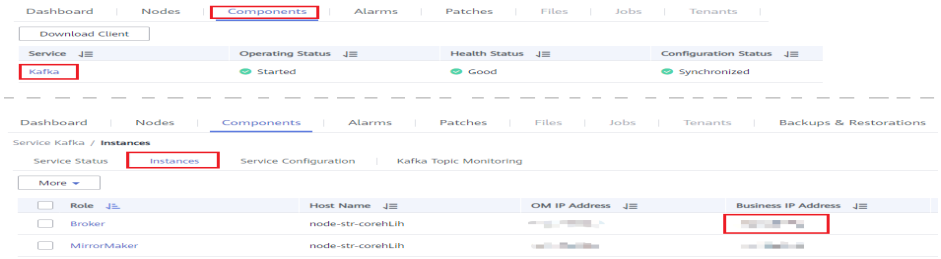
kafka-topics.sh --create --zookeeper <*IP address of the node where the ZooKeeper instance resides***:2181/kafka> --partitions 2 --replication-factor 2 --topic <***Topic name***>**

----End

Managing Messages in Kafka Topics

Step 1 Click the **Components** tab. On this tab page, select **Kafka**. On the page that is displayed, click the **Instances** tab. On the **Instances** tab page, view the IP addresses of Kafka instances. Record the IP address of any Kafka instance, for example, **192.168.7.15**.

 Copyright © Huawei Technologies Co., Ltd.



Step 2 Log in to the master node and run the following command to generate messages in a topic test:

kafka-console-producer.sh --broker-list <*IP address of the node where the Kafka instance resides***:9092> --topic <***Topic name***> --producer.config /opt/hadoopclient/Kafka/kafka/config/producer.properties**

Enter the specified content as the messages generated by the producer and press **Enter** to send the messages. To stop generating messages, press **Ctrl+C** to exit.

Step 3 Consume messages in the topic test.

kafka-console-consumer.sh --topic <*Topic name***> --bootstrap-server <***IP address of the node where the Kafka instance resides***:9092> --consumer.config /opt/hadoopclient/Kafka/kafka/config/consumer.properties**

----End

 Copyright © Huawei Technologies Co., Ltd.

6 Using HBase from Scratch

MapReduce Service (MRS) provides enterprise-level big data clusters on the cloud. Tenants can fully control the clusters and run big data components such as Hadoop, Spark, HBase, and Kafka in the clusters.

This section uses a cluster with Kerberos authentication disabled as an example to describe how to log in to the HBase client, create a table, insert data into the table, and modify the table.

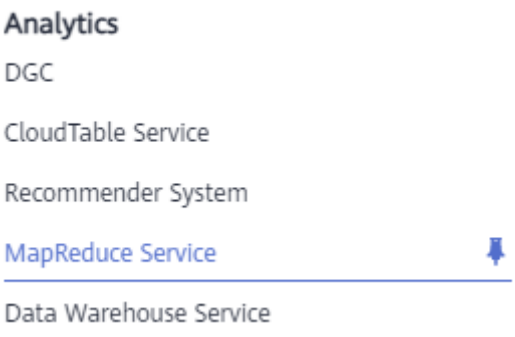
You can get started by reading the following topics:

1. **Preparing an MRS Cluster**
2. **Installing the HBase Client**
3. **Creating a Table Using the HBase Client**

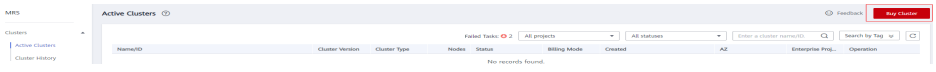
Preparing an MRS Cluster

Step 1 Purchase an MRS cluster.

1. Log in to the Huawei Cloud management console.
2. Choose **Analytics** > **MapReduce Service** to go to the MRS console.



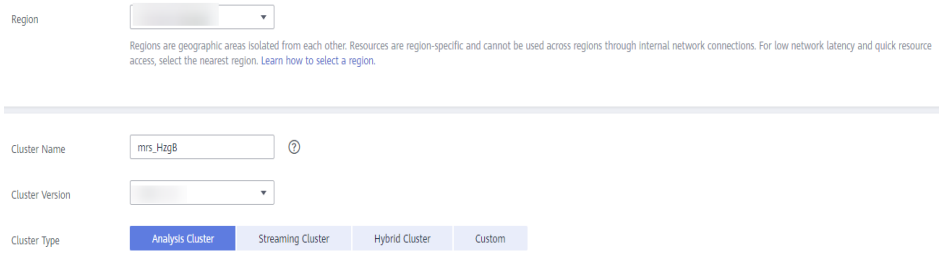
3. Choose **Clusters** > **Active Clusters** and click **Buy Cluster**. On the displayed page, click **Custom Config**.



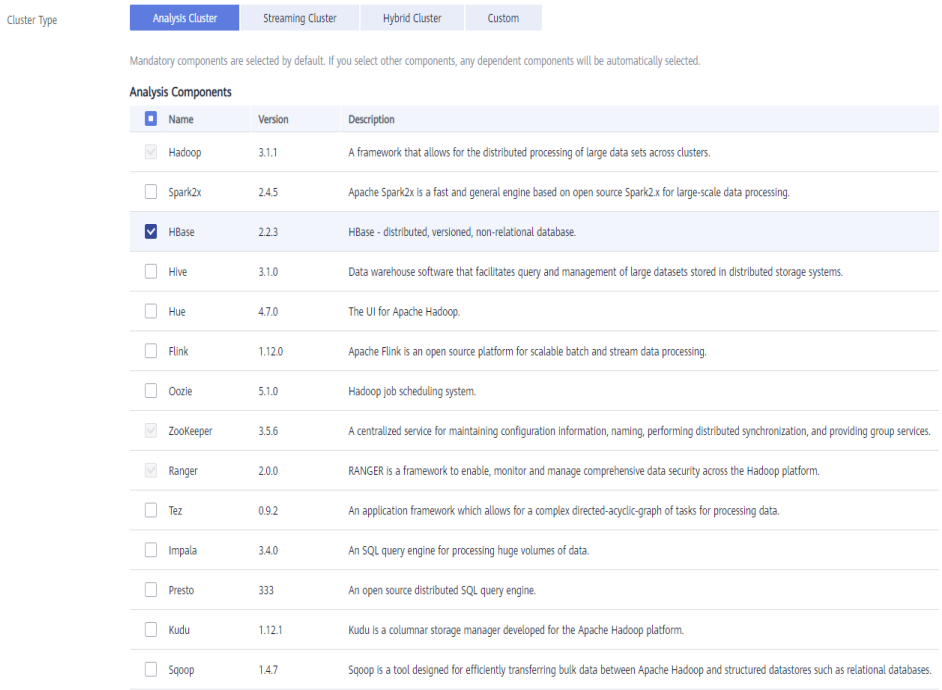
Step 2 Set the following parameters and click **Next**.

 Copyright © Huawei Technologies Co., Ltd.

- **Region**: Select a region as required.
- **Cluster Name**: Enter **mrs_demo** or specify a name according to naming rules.
- **Cluster Version**: Select **MRS 3.1.0**.



- **Cluster Type**: Select **Analysis Cluster** and select HBase.



Step 3 On the **Configure Hardware** page, set the parameters by referring to **Table 6-1**, and click **Next**.

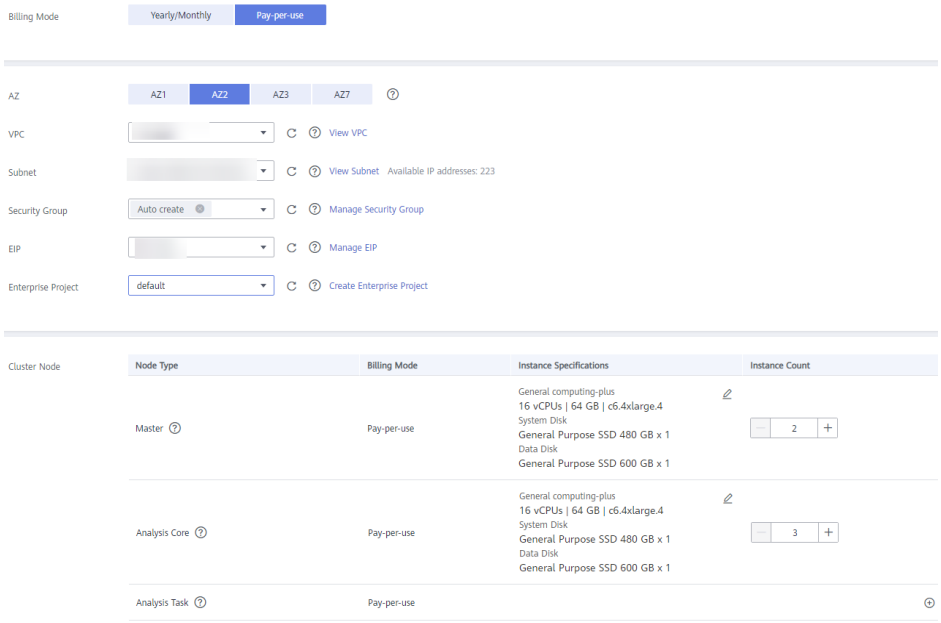
Table 6-1 MRS cluster hardware configuration

Parameter	Example Value
Billing Mode	Pay-per-use
AZ	AZ2
VPC	Retain the default value. You can also click **View VPC** to create a VPC.
EIP	You can select an existing EIP from the drop-down list. If no EIP is available in the drop-down list, click **Manage EIP** to access the **EIPs** page to create one.

 Copyright © Huawei Technologies Co., Ltd.

Parameter	Example Value
Enterprise Project	default

Figure 6-1 Hardware configurations



Step 4 Configure advanced options.

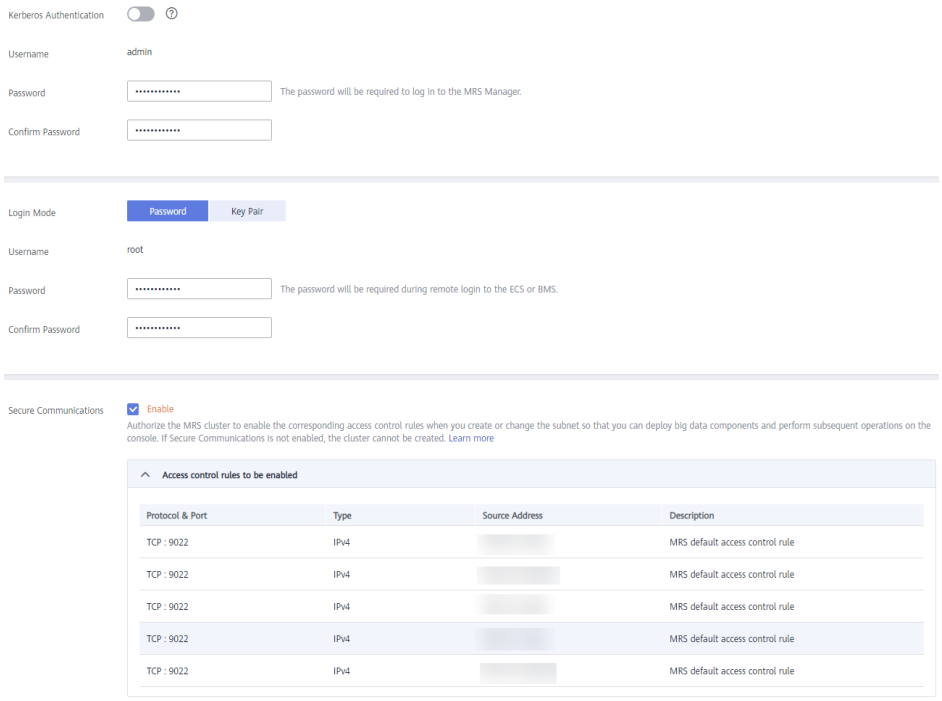
1. On the **Set Advanced Options** page, set parameters by referring to **Table 6-2**.

   Table 6-2 MRS cluster advanced options

Parameter	Example Value
Kerberos Authentication	Disabled
Password	Test@!123456
Confirm Password	Test@!123456
Login Mode	Password
Password	Test@#123456
Confirm Password	Test@#123456
Secure Communications	Select **Enable**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 6-2 Advanced options



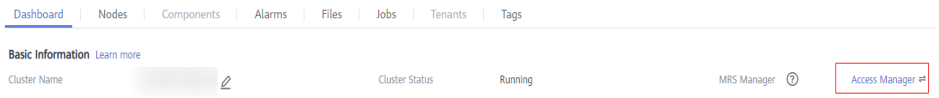
2. Click **Buy Now**. The page is displayed showing that the task has been submitted.
3. Click **Back to Cluster List**. You can view the status of the cluster on the **Active Clusters** page.
4. Wait for the cluster creation to complete. The initial status of the cluster is **Starting**. After the cluster has been created successfully, the cluster status becomes **Running**.

----End

Installing the HBase Client

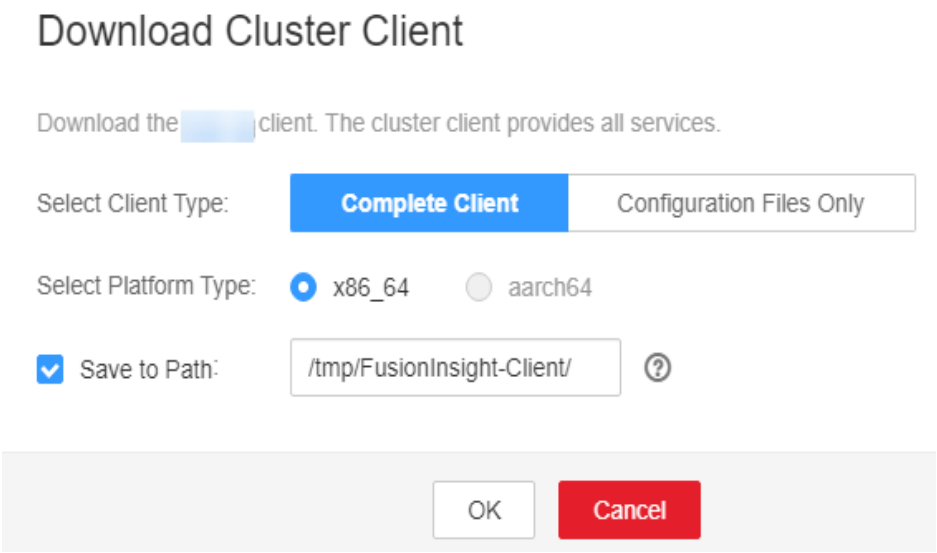
Step 1 Choose **Clusters** > **Active Clusters** and click **mrs_demo**. The cluster information page is displayed.

Step 2 Click **Access Manager** next to **MRS Manager**. On the displayed page, configure the EIP information and click **OK**. Enter the username and password to access FusionInsight Manager.



Step 3 Choose **Cluster** > **Services** > **HBase**, and select **Download Client** from the **More** drop-down list. Select **Complete Client**, the corresponding platform type, and **Save to path**, and click OK.

 Copyright © Huawei Technologies Co., Ltd.



Step 4 Log in to the active node as the **root** user.

Step 5 Go to the directory where the installation package is stored and run the following commands to decompress and verify the installation package, and decompress the obtained installation file:

cd /tmp/FusionInsight-Client

tar -xvf FusionInsight_Cluster_1_HBase_Client.tar

sha256sum -c FusionInsight_Cluster_1_HBase_ClientConfig.tar.sha256

tar -xvf FusionInsight_Cluster_1_HBase_ClientConfig.tar

Step 6 Go to the directory where the installation package is stored, and run the following command to install the client to a specified directory (an absolute path), for example, **/opt/hbaseclient**:

cd /tmp/FusionInsight-Client/FusionInsight_Cluster_1_HBase_ClientConfig

Run the **./install.sh /opt/hbaseclient** command and wait until the client installation is complete.

Step 7 Check whether the client is successfully installed.

cd /opt/hbaseclient

source bigdata_env

hbase shell

If the command is successfully executed, the HBase client is successfully installed.

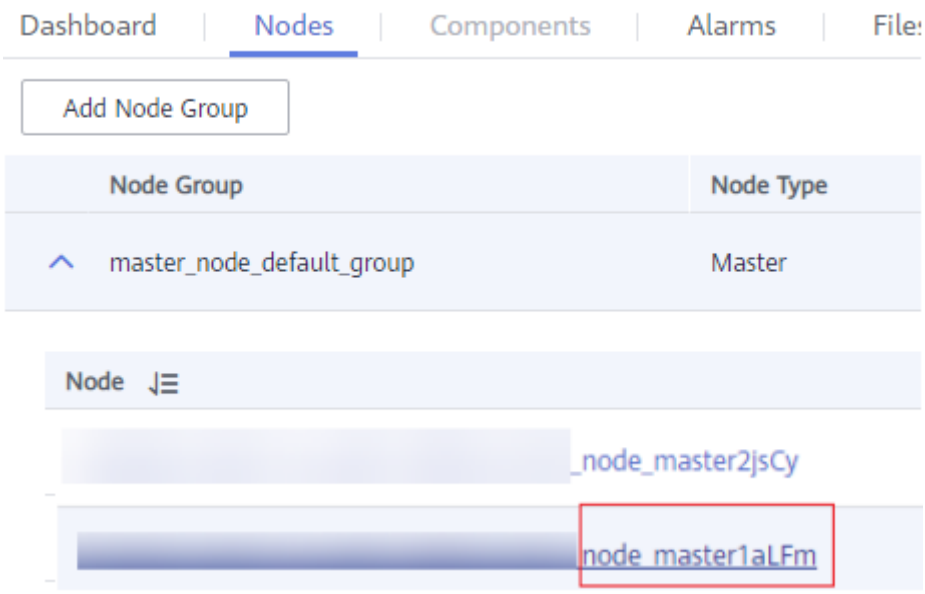
----End

Creating a Table Using the HBase Client

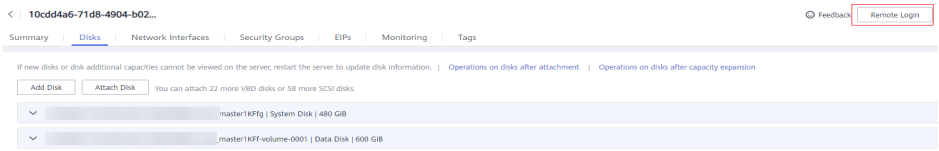
Step 1 Log in to the master node using VNC.

1. On the MRS console, choose **Clusters** > **Active Clusters**, and select **mrs_demo** from the cluster list. Click **Nodes**, and click the node whose name contains **master1** to access its ECS details page.

 Copyright © Huawei Technologies Co., Ltd.



2. Click **Remote Login** in the upper right corner of the page to log in to the master node as user **root**. The password is the one set when the cluster is purchased.



Step 2 Run the following command to go to the client directory:

cd /opt/hbaseclient

Step 3 Run the following command to configure environment variables:

source bigdata_env

NOTE

If Kerberos authentication is enabled for the cluster, run the following command to authenticate the current user. The current user must have the permission to create HBase tables.

For example:

kinit *hbaseuser*

Step 4 Run the following command to access the HBase shell CLI:

hbase shell

Step 5 Run the HBase client command to create the **user_info** table.

1. Create the **user_info** table and add related data.

   create '*user_info*',{**NAME** => 'i'}

   put '*user_info*','*12005000201*','***i:name***','*A*'

   put '*user_info*','*12005000201*','***i:gender***','*Male*'

   put '*user_info*','*12005000201*','***i:age***','*19*'

 Copyright © Huawei Technologies Co., Ltd.

put '*user_info*','*12005000201*','***i:address***','*City A*'

2. Add users' educational backgrounds and professional titles to the **user_info** table.

   put '*user_info*','*12005000201*','**i:degree**','*master*'

   put '*user_info*','*12005000201*','**i:pose**','*manager*'

3. Query user names and addresses by user ID.

   scan'*user_info*',
   {**STARTROW**=>'*12005000201*',**STOPROW**=>'*12005000201*',**COLUMNS**=>['**i:name**','**i:address**']}

   ```
   ROW                                   COLUMN
   +CELL
    12005000201                          column=i:address, timestamp=2021-10-30T10:21:42.196, value=City
   A
    12005000201                          column=i:name, timestamp=2021-10-30T10:21:18.594,
   value=A
   1 row(s)
   Took 0.0996 seconds
   ```

4. Query information by user name.

   scan'*user_info*',{**FILTER**=>"SingleColumnValueFilter('i','name',=,'binary:*A*')"}

   ```
   ROW                                   COLUMN
   +CELL
    12005000201                          column=i:address, timestamp=2021-10-30T10:21:42.196, value=City
   A
    12005000201                          column=i:age, timestamp=2021-10-30T10:21:30.777,
   value=19
    12005000201                          column=i:degree, timestamp=2021-10-30T10:21:53.284,
   value=master
    12005000201                          column=i:gender, timestamp=2021-10-30T10:21:18.711,
   value=Male
    12005000201                          column=i:name, timestamp=2021-10-30T10:21:18.594,
   value=A
    12005000201                          column=i:pose, timestamp=2021-10-30T10:22:07.152,
   value=manager
   1 row(s)
   Took 0.2158 seconds
   ```

5. Delete user data from the user information table.

   delete'*user_info*','*12005000201*','i'

6. Delete the user information table.

   disable *'user_info'*

   drop *'user_info'*

----End

 Copyright © Huawei Technologies Co., Ltd.

7 Modifying MRS Configurations

After an MRS cluster is created, you can modify configuration parameters of services in the cluster on the MRS console or Manager.

This section uses the **hbase.log.maxbackupindex** parameter of the HBase service as an example to describe how to modify the MRS configuration parameters.

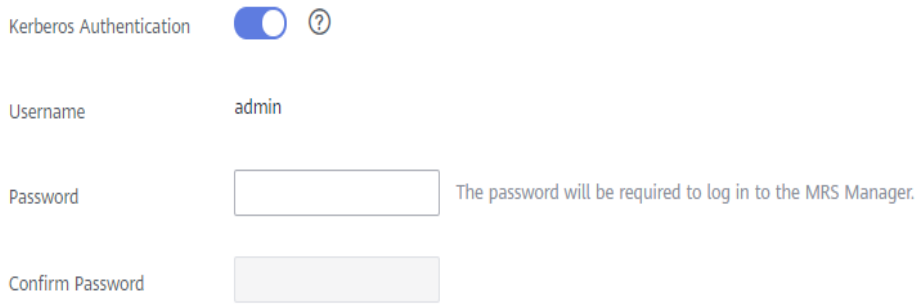
You can get started by reading the following topics:

1. **Modifying Service Parameters on the MRS Console**
2. **Modifying Service Parameters on FusionInsight Manager**

Modifying Service Parameters on the MRS Console

Step 1 Create a security cluster. For details, see **Custom Purchase of a Cluster**. Enable **Kerberos Authentication**, configure **Password**, and confirm the password. This password is used to log in to Manager. Keep it secure.

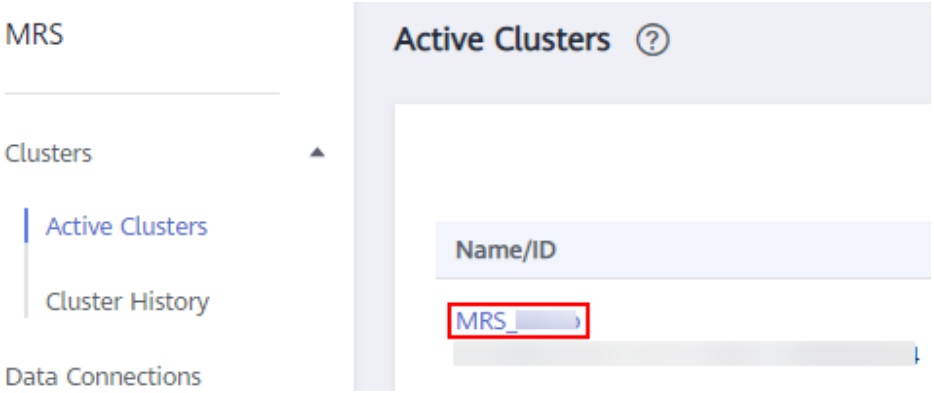
Figure 7-1 Setting security cluster parameters



Step 2 Log in to the MRS console. In the navigation pane on the left, choose **Clusters** > **Active Clusters** and click a cluster name.

 Copyright © Huawei Technologies Co., Ltd.

Figure 7-2 Clicking a cluster name

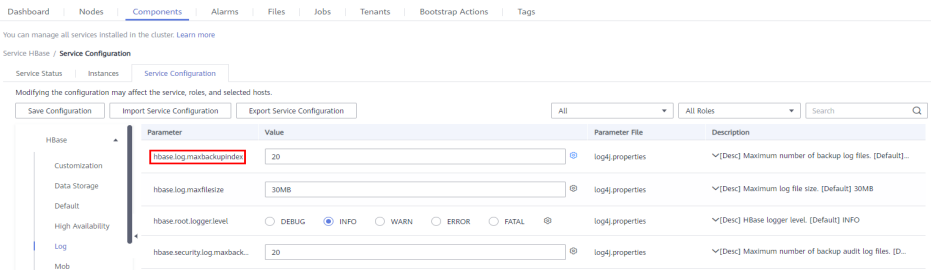


Step 3 Choose **Components** > **HBase**, click **Service Configuration**, and choose **All** in the upper right corner of the page.

Step 4 In the navigation tree on the left, choose **HBase** > **Log**.

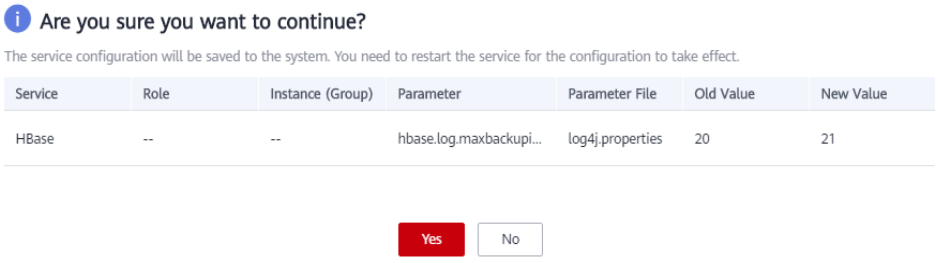
Step 5 Locate the **hbase.log.maxbackupindex** parameter and change its value based on service requirements.

Figure 7-3 Changing the parameter value



Step 6 Click **Save Configuration**. In the displayed dialog box, confirm the changed parameter value, and click **Yes**. Wait for the system to save and update the configuration, and click **Finish**.

Figure 7-4 Confirming the modification

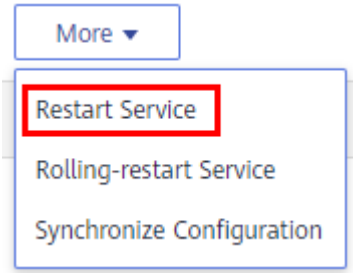


Step 7 Check the current service configuration status.

Click **Service Status** to view the current service configuration status. If the configuration of a service has expired, click **More** and select **Restart Service** to restart the service. In the displayed dialog box, click **Yes**. Then wait until the service is restarted.

 Copyright © Huawei Technologies Co., Ltd.

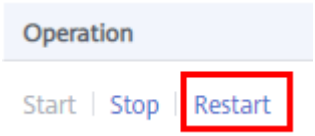
Figure 7-5 Restarting a service



Step 8 Check the service configuration status of related services.

Return to the **Components** page to check the configuration status of related services. If the configuration of a service has expired, click **Restart** in the **Operation** column of the service. In the displayed dialog box, click **Yes** to restart it.

Figure 7-6 Restarting a service



----End

Modifying Service Parameters on FusionInsight Manager

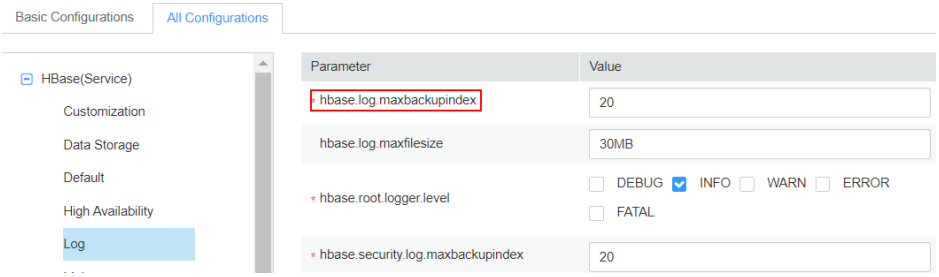
Step 1 Create a cluster and log in to FusionInsight Manager. For details, see **Creating a Security Cluster and Logging In to Manager**.

Step 2 Choose **Cluster** > **Services** > **HBase**, choose **Configurations**, and click **All Configurations**.

Step 3 Choose **HBase(Service)** > **Log**.

Step 4 Locate the **hbase.log.maxbackupindex** parameter and change its value based on service requirements.

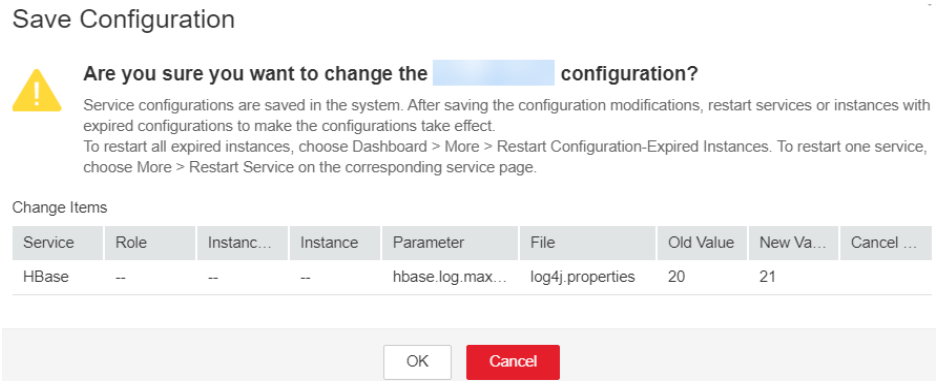
Figure 7-7 Changing the parameter value



Step 5 Click **Save**. In the displayed dialog box, confirm the changed parameter value and click **OK**. Wait for the system to save and update the configuration, and click **Finish**.

 Copyright © Huawei Technologies Co., Ltd.

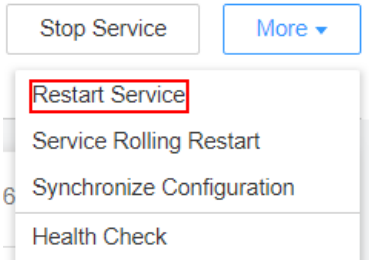
Figure 7-8 Confirming the modification



Step 6 Check the current service configuration status.

Click **Dashboard** to view the current service configuration status. If the configuration of a service has expired, click **More** and select **Restart Service**. Then enter the password and click **OK** to restart the service. Wait until the service is restarted.

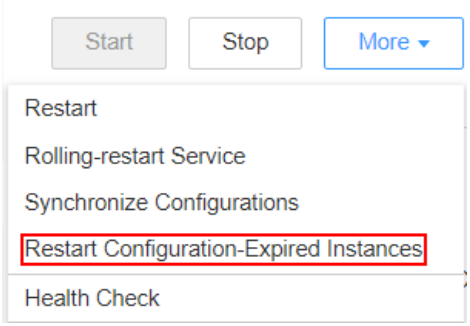
Figure 7-9 Restarting a service



Step 7 Check the service configuration status of related services.

Choose **Cluster** > **Service** to view the configuration status of other related services. If the configuration of a service has expired, choose **Cluster** > **Dashboard**, select **Restart Configuration-Expired Instances** from the **More** drop-down list, enter the password, and click **OK** to restart it.

Figure 7-10 Restarting configuration-expired instances



----End

 Copyright © Huawei Technologies Co., Ltd.

8 Configuring Auto Scaling for an MRS Cluster

In big data application scenarios, especially real-time data analysis and processing, the number of cluster nodes needs to be dynamically adjusted according to data volume changes to provide proper resources. The auto scaling function of MRS enables clusters to be automatically scaled out or in based on cluster load.

- Auto scaling rules: You can increase or decrease Task nodes based on real-time cluster loads. Auto scaling will be triggered when the data volume changes but there may be some delays.
- Resource plan (setting the task node quantity based on the time range): If the data volume changes periodically, you can create resource plans to resize the cluster before the data volume changes, thereby avoiding delays in increasing or decreasing resources.

You can configure either auto scaling rules or resource plans or both of them to trigger the auto scaling. This section describes how to configure auto scaling rules for MRS clusters based on service scenarios.

You can get started by reading the following topics:

1. **Creating a Cluster and Configuring Task Nodes**
2. **Scenario 1: Using Auto Scaling Rules Alone**
3. **Scenario 2: Using Resource Plans Alone**
4. **Scenario 3: Using Auto Scaling Rules and Resource Plans Together**

Creating a Cluster and Configuring Task Nodes

NOTE

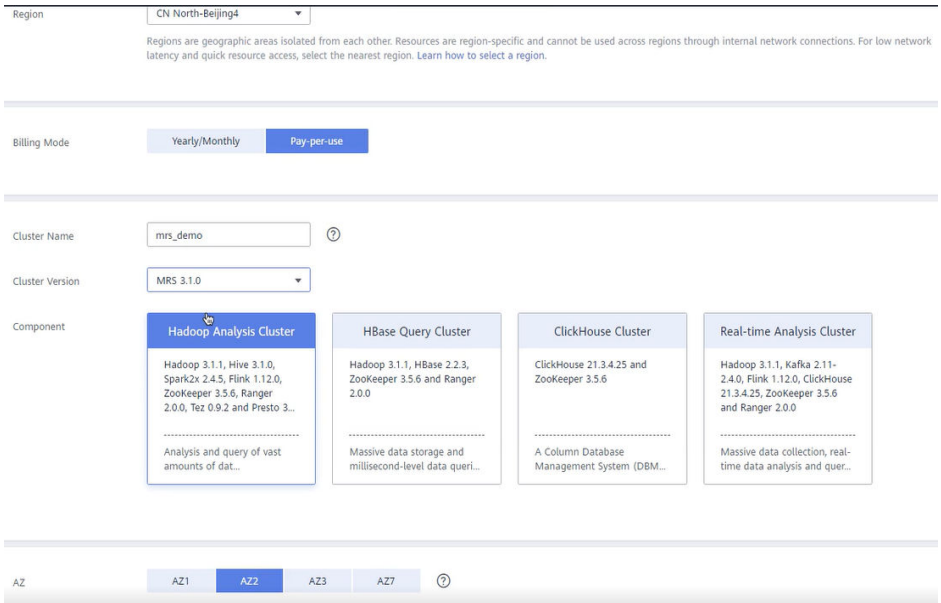
- The following steps use a Hadoop analysis cluster of MRS 3.1.0 as an example to describe how to quickly purchase a cluster.
- Only task node groups support auto scaling. Check whether a task node exists in the current cluster before configuring auto scaling.

Step 1 Log in to the Huawei Cloud management console and choose **Analytics** > **MapReduce Service**. Click **Buy Cluster**, configure the parameters on the **Quick Config** tab page, and click **Buy Now**.

 Copyright © Huawei Technologies Co., Ltd.

Table 8-1 Parameters (for reference only)

Parameter	Value
Region	Select the region based on service requirements.
Billing Mode	Pay-per-use
Cluster Name	MRS_demo
Cluster Version	MRS 3.1.0
Component	Hadoop Analysis Cluster
AZ	AZ2
VPC	vpc-gggg
Subnet	subnet-64db
Enterprise Project	default
Kerberos Authentication	Disabled
Username	root/admin
Password	Set the password for logging in to the cluster management page and ECS node, for example, **Test!@12345**.
Confirm Password	Enter the password again.
Secure Communications	Select **Enable**.



Step 2 Click the created cluster and click **Nodes** to check whether there is a Task node in the cluster.

 Copyright © Huawei Technologies Co., Ltd.

- If yes, no further action is required.
- If no, go to **Step 3**.

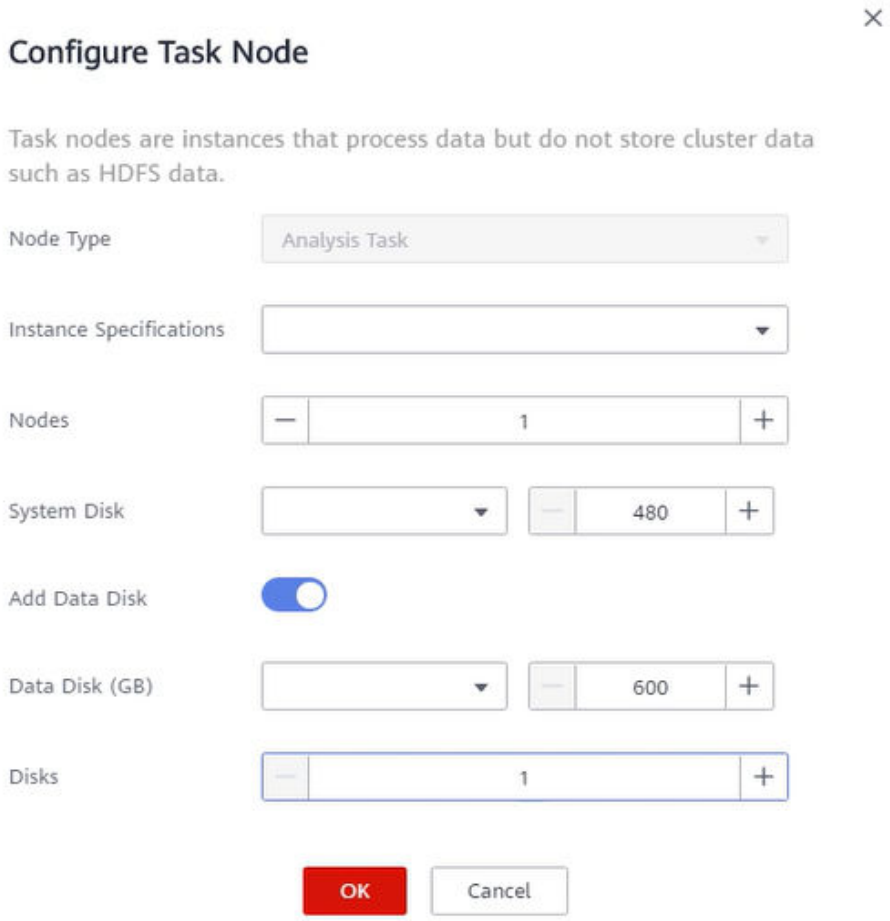
Step 3 Configure a task node.

1. On the **Nodes** page, click **Configure Task Node**.

   NOTE

   For MRS 3.*x* or later, **Configure Task Node** applies only to analysis clusters, streaming clusters, and hybrid clusters.

2. Set required parameters.



3. Click **OK**.

----End

Scenario 1: Using Auto Scaling Rules Alone

The following is an example of the service scenario:

The number of nodes needs to be dynamically adjusted based on the Yarn resource usage. When the memory available for Yarn is less than 20%, five nodes need to be added. When the memory available for Yarn is greater than 70%, five nodes need to be reduced. The number of nodes in a task node group ranges from 1 to 10.

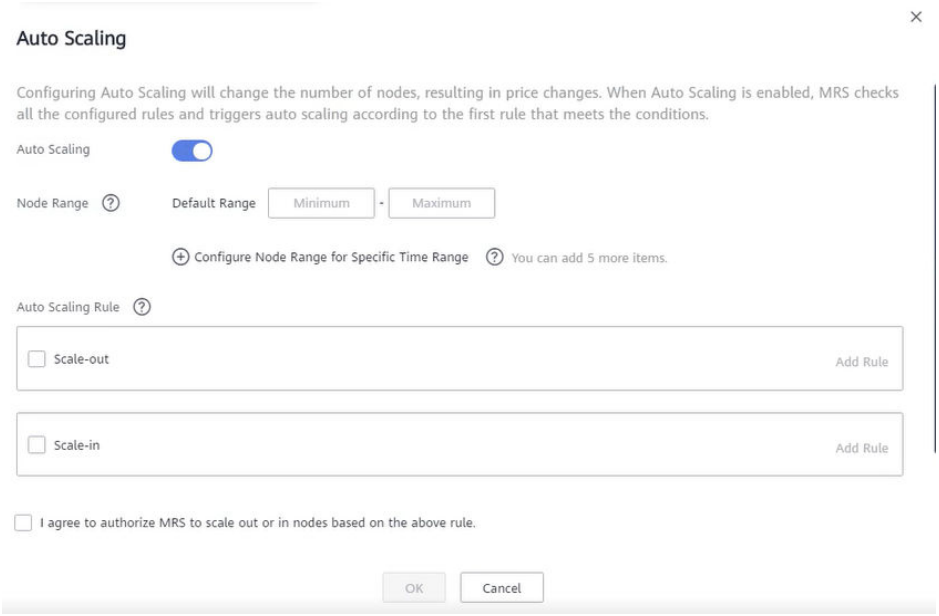
 Copyright © Huawei Technologies Co., Ltd.

Step 1 Create a cluster and configure task nodes by referring to **Creating a Cluster and Configuring Task Nodes**.

Step 2 On the MRS console, choose **Clusters** > **Active Clusters** and click the name of the target cluster to access its details page.

Step 3 Click **Nodes** and click **Auto Scaling** in the **Operation** column of the task node group.

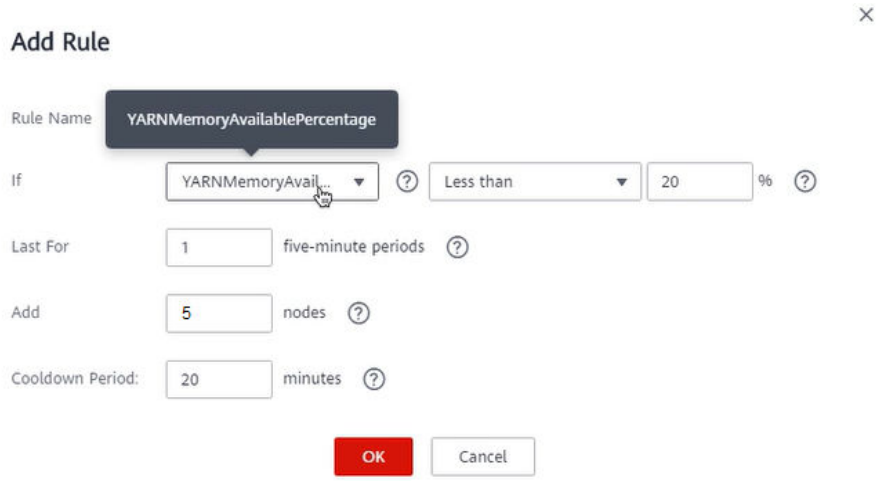
Step 4 On the **Auto Scaling** page, click to enable auto scaling and set **Node Range** to **1-10**.



Step 5 Configure scale-out rules.

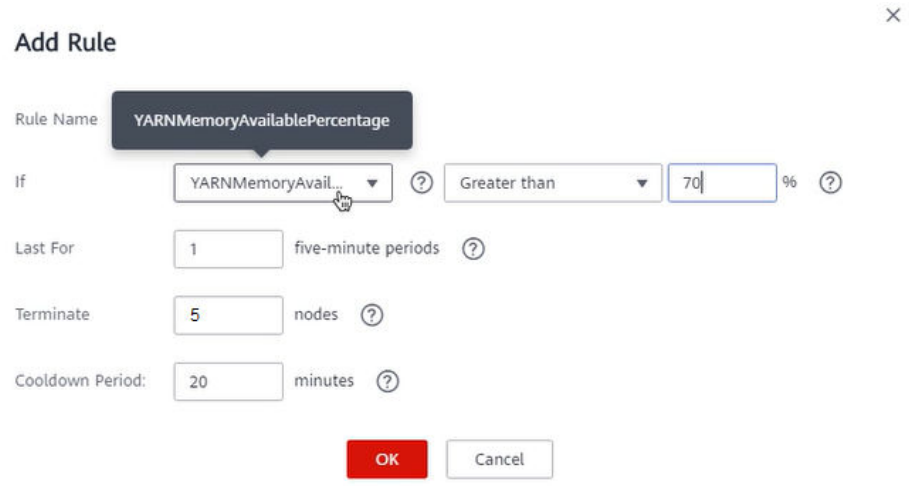
1. Select **Scale-out** in the **Auto Scaling Rule** area.
2. Click **Add Rule** next to **Scale-out**.
3. On the **Add Rule** page, set the related parameters.
   - **Rule Name**: Retain the default value, for example, **default-expand-2**.
   - **If**: Select **YARNMemoryAvailablePercentage** and **Less than** from the drop-down lists, and set the percentage to 20% (for details about related metrics, see **Table 8-2**).
   - **Last For**: Set it to **1 five-minute periods**.
   - **Add**: Set it to **5 nodes**.
   - **Cooldown Period**: Set it to **20 minutes**.
4. Click **OK**.

 Copyright © Huawei Technologies Co., Ltd.



Step 6 Configure scale-in rules.

1. Select **Scale-in** in the **Auto Scaling Rule** area.
2. Click **Add Rule** next to **Scale-in**.
3. On the **Add Rule** page, set the related parameters.
   - **Rule Name**: Retain the default value, for example, **default-shrink-2**.
   - **If**: Select **YARNMemoryAvailablePercentage** and **Greater than** from the drop-down lists, and set the percentage to 70% (for details about related indicators, see **Table 8-2**).
   - **Last For**: Set it to **1 five-minute periods**.
   - **Terminate**: Set it to **5 nodes**.
   - **Cooldown Period**: Set it to **20 minutes**.
4. Click **OK**.



Step 7 Select **I agree to authorize MRS to scale out or in nodes based on the above rule**.

Step 8 Click **OK**.

----End

 Copyright © Huawei Technologies Co., Ltd.

Scenario 2: Using Resource Plans Alone

The following is an example of the service scenario:

The number of nodes needs to be increased or decreased periodically. 10 nodes are required from 10:00 to 22:00 due to heavy service traffic, and five nodes are required in other time segments.

Step 1 Create a cluster and configure task nodes by referring to **Creating a Cluster and Configuring Task Nodes**.

Step 2 On the MRS console, choose **Clusters** > **Active Clusters** and click the name of the target cluster to access its details page.

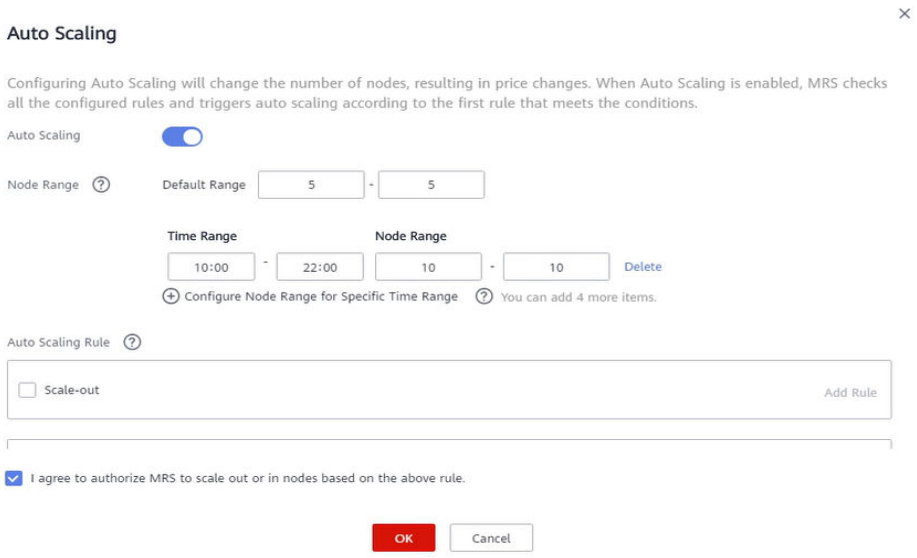
Step 3 Click **Nodes** and click **Auto Scaling** in the **Operation** column of the task node group.

Step 4 On the **Auto Scaling** page, enable auto scaling and configure **Node Range**.

- **Auto Scaling**: Enable.
- **Node Range**: 5-5.

Step 5 Click **Configure Node Range for Specified Time Range** under **Default Range** and set related parameters.

- **Time Period**: 10:00-22:00.
- **Node Range**: 10-10.



Step 6 Select **I agree to authorize MRS to scale out or in nodes based on the above rule**.

Step 7 Click **OK**.

----End

Scenario 3: Using Auto Scaling Rules and Resource Plans Together

The following is an example of configuring both auto scaling rules and resource plans.

 Copyright © Huawei Technologies Co., Ltd.

The following is an example of the service scenario:

Even though the service data volume for real-time processing changes regularly from 7:00 to 13:00 every day, it is still unstable. Assume that during 7:00 to 13:00, the number of required task nodes ranges from 5 to 8, and in other time ranges, the number of required task nodes ranges from 2 to 4 based on the number of tasks running on Yarn.

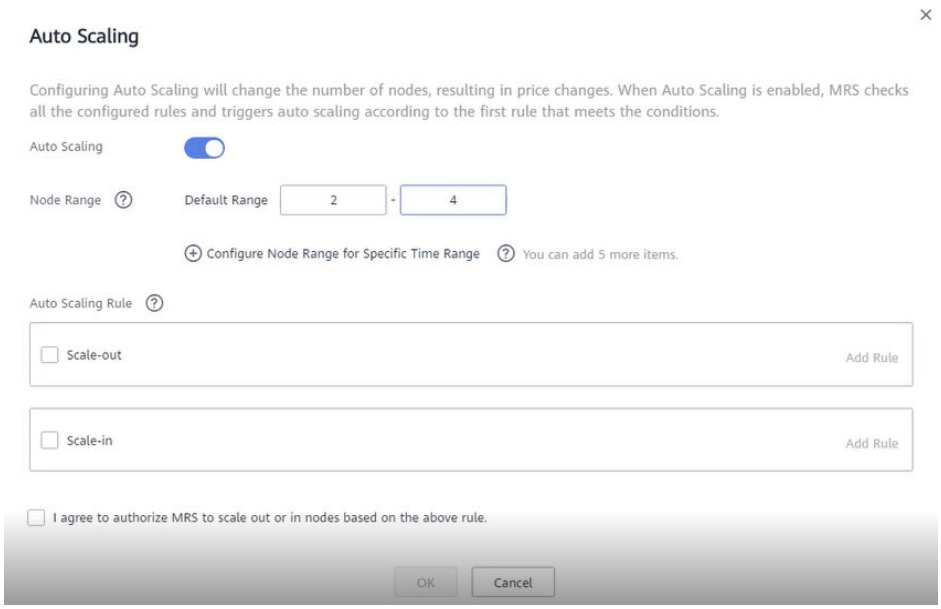
Step 1 Create a cluster and configure task nodes by referring to **Creating a Cluster and Configuring Task Nodes**.

Step 2 On the MRS console, choose **Clusters** > **Active Clusters** and click the name of the target cluster to access its details page.

Step 3 Click **Nodes** and click **Auto Scaling** in the **Operation** column of the task node group.

Step 4 On the **Auto Scaling** page, enable auto scaling, and set **Node Range** to 2-4.

Figure 8-1 Configuring auto scaling

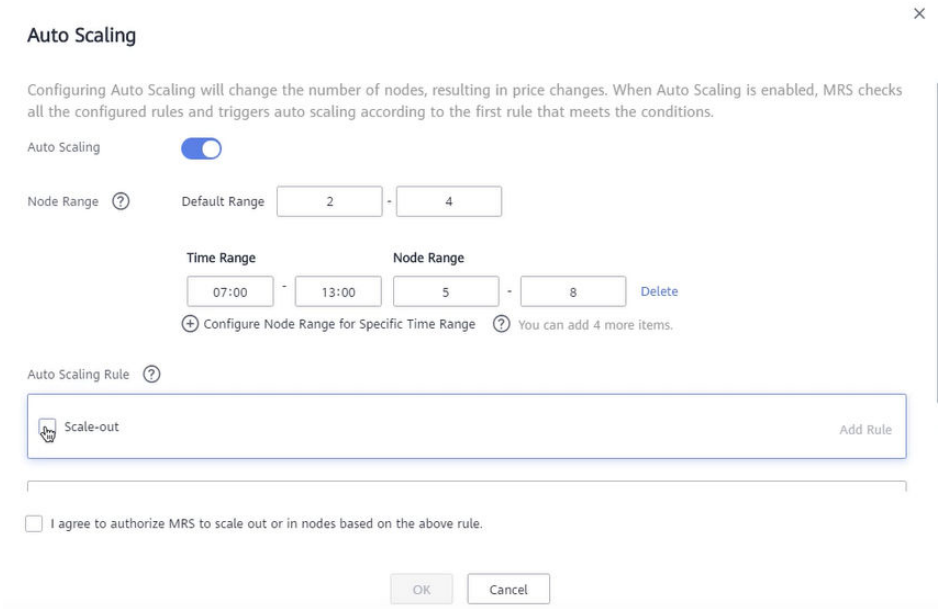


Step 5 Configure a resource plan.

1. Click **Configure Node Range for Specific Time Range** under **Default Range**.
2. Configure the **Time Range** and **Node Range** parameters.

 Copyright © Huawei Technologies Co., Ltd.

Figure 8-2 Auto scaling



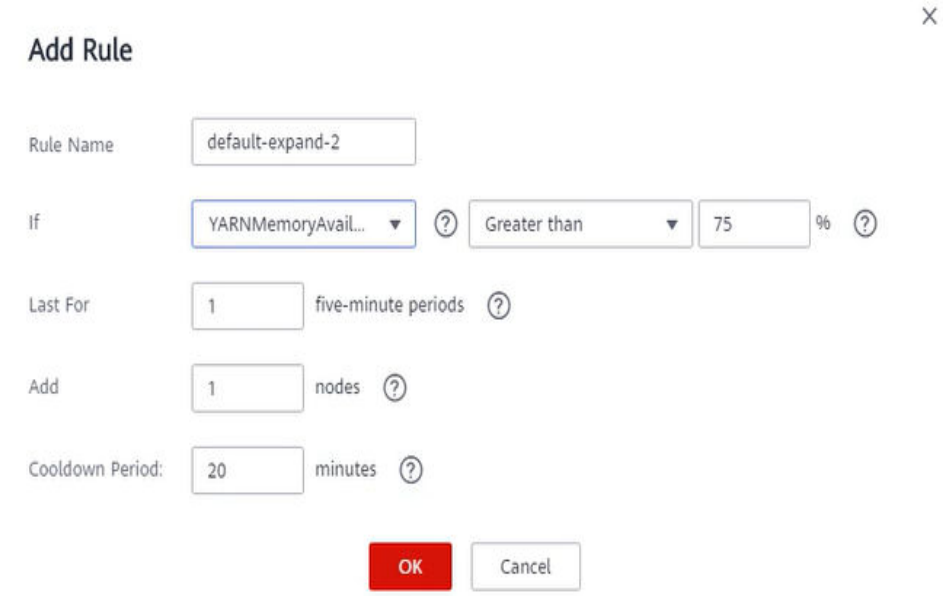
Time Range: Set it to **07:00-13:00**.

Node Range: Set it to **5-8**.

Step 6 Configure an auto scaling rule.

1. Select **Scale-out**.
2. Click **Add Rule** on the right.

Figure 8-3 Adding a rule



Rule Name: **default-expand-2**.

If: Select the rule objects and constraints from the drop-down list boxes, for example, YARNAppRunning is greater than 75.

 Copyright © Huawei Technologies Co., Ltd.

Last For: Set it to **1 five-minute periods**.

Add: Set it to **1 node**.

Cooldown Period: Set it to **20 minutes**.

3. Click **OK**.

Step 7 Select **I agree to authorize MRS to scale out or in nodes based on the above rule**.

Step 8 Click **OK**.

----End

Reference Information

When adding a rule, you can refer to **Table 8-2** to configure the corresponding metrics.

NOTE

- Hybrid clusters support all metrics of analysis and streaming clusters.
- The accuracy of different value types in **Table 8-2** is as follows:
  - **Integer**: integer
  - **Percentage**: 0.01
  - **Ratio**: 0.01

Table 8-2 Auto scaling metrics

Cluster Type	Metric	Value Type	Description
Streaming cluster	StormSlotAvailable	Integer	Number of available Storm slots. Value range: 0 to 2147483646.
	StormSlotAvailablePercentage	Percentage	Percentage of available Storm slots, that is, the proportion of the available slots to total slots. Value range: 0 to 100.
	StormSlotUsed	Integer	Number of used Storm slots. Value range: 0 to 2147483646.
	StormSlotUsedPercentage	Percentage	Percentage of the used Storm slots, that is, the proportion of the used slots to total slots. Value range: 0 to 100.
	StormSupervisorMemAverageUsage	Integer	Average memory usage of the Supervisor process of Storm. Value range: 0 to 2147483646.

 Copyright © Huawei Technologies Co., Ltd.

Cluster Type	Metric	Value Type	Description
	StormSupervisor-MemAverageUsagePercentage	Percentage	Average percentage of the used memory of the Supervisor process of Storm to the total memory of the system. Value range: 0 to 100.
	StormSupervisorCPUAverageUsagePercentage	Percentage	Average percentage of the used CPUs of the Supervisor process of Storm to the total CPUs. Value range: 0 to 6,000.
Analysis cluster	YARNAppPending	Integer	Number of pending tasks on Yarn. Value range: 0 to 2147483646.
	YARNAppPendingRatio	Ratio	Ratio of pending tasks on Yarn, that is, the ratio of pending tasks to running tasks on Yarn. Value range: 0 to 2147483646.
	YARNAppRunning	Integer	Number of running tasks on Yarn. Value range: 0 to 2147483646.
	YARNContainerAllocated	Integer	Number of containers allocated to YARN. Value range: 0 to 2147483646.
	YARNContainerPending	Integer	Number of pending containers on Yarn. Value range: 0 to 2147483646.
	YARNContainerPendingRatio	Ratio	Ratio of pending containers on Yarn, that is, the ratio of pending containers to running containers on Yarn. Value range: 0 to 2147483646.
	YARNCPUAllocated	Integer	Number of virtual CPUs (vCPUs) allocated to Yarn. Value range: 0 to 2147483646.
	YARNCPUAvailable	Integer	Number of available vCPUs on Yarn. Value range: 0 to 2147483646.
	YARNCPUAvailablePercentage	Percentage	Percentage of available vCPUs on Yarn, that is, the proportion of available vCPUs to total vCPUs. Value range: 0 to 100.

 Copyright © Huawei Technologies Co., Ltd.

Cluster Type	Metric	Value Type	Description
	YARNCPUPending	Integer	Number of pending vCPUs on Yarn. Value range: 0 to 2147483646.
	YARNMemoryAllocated	Integer	Memory allocated to Yarn. The unit is MB. Value range: 0 to 2147483646.
	YARNMemoryAvailable	Integer	Available memory on Yarn. The unit is MB. Value range: 0 to 2147483646.
	YARNMemoryAvailablePercentage	Percentage	Percentage of available memory on Yarn, that is, the proportion of available memory to total memory on Yarn. Value range: 0 to 100.
	YARNMemoryPending	Integer	Pending memory on Yarn. Value range: 0 to 2147483646.

When adding a resource plan, you can set parameters by referring to **Table 8-3**.

Table 8-3 Configuration items of a resource plan

Parameter	Description
Time range	Start time and end time of a resource plan are accurate to minutes, with the value ranging from **00:00** to **23:59**. For example, if a resource plan starts at 8:00 and ends at 10:00, set this parameter to **8:00-10:00**. The end time must be at least 30 minutes later than the start time.
Node range	The number of nodes in a resource plan ranges from **0** to **500**. In the time range specified in the resource plan, if the number of task nodes is less than the specified minimum number of nodes, it will be increased to the specified minimum value of the node range at a time. If the number of task nodes is greater than the maximum number of nodes specified in the resource plan, the auto scaling function reduces the number of task nodes to the maximum value of the node range at a time. The minimum number of nodes must be less than or equal to the maximum number of nodes.

 Copyright © Huawei Technologies Co., Ltd.

9 Configuring Hive with Storage and Compute Decoupled

MRS allows you to store data in OBS and use an MRS cluster for data computing only. In this way, storage and compute are decoupled. You can use the IAM service to perform simple configurations to access OBS.

This section describes how to create a Hive table to store data to OBS.

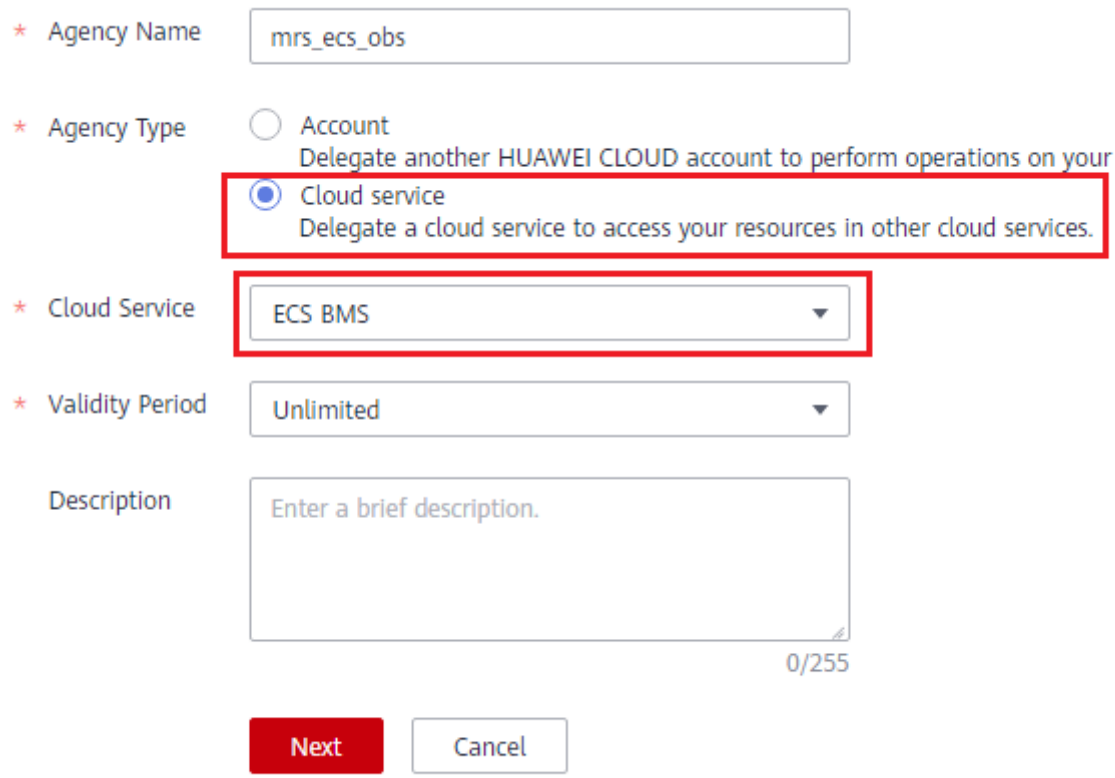
1. **Creating an ECS Agency**
2. **Configuring an Agency for an MRS Cluster**
3. **Creating an OBS File System**
4. **Accessing the OBS File System Through Hive**

Creating an ECS Agency

1. Log in to the Huawei Cloud management console.
2. Choose **Service List** > **Management & Governance** > **Identity and Access Management**.
3. Click **Agencies**. On the displayed page, click **Create Agency**.
4. Enter an agency name, for example, **mrs_ecs_obs**.
5. Set **Agency Type** to **Cloud service** and select **ECS BMS** to authorize ECS or BMS to invoke OBS.
6. Set **Validity Period** to **Unlimited** and click **Next**.

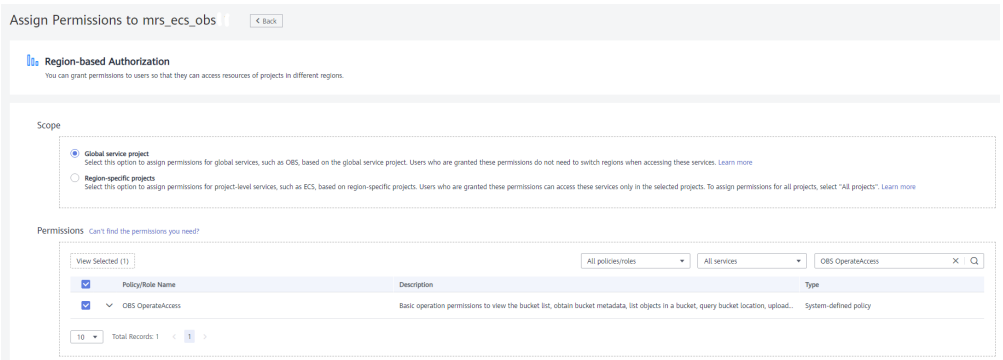
 Copyright © Huawei Technologies Co., Ltd.

Figure 9-1 Creating an agency



7. On the page that is displayed, select **Global service project**, search for the **OBS OperateAccess** policy, and select the **OBS OperateAccess** policy.

Figure 9-2 Assigning permissions



8. Click **OK**.

Configuring an Agency for an MRS Cluster

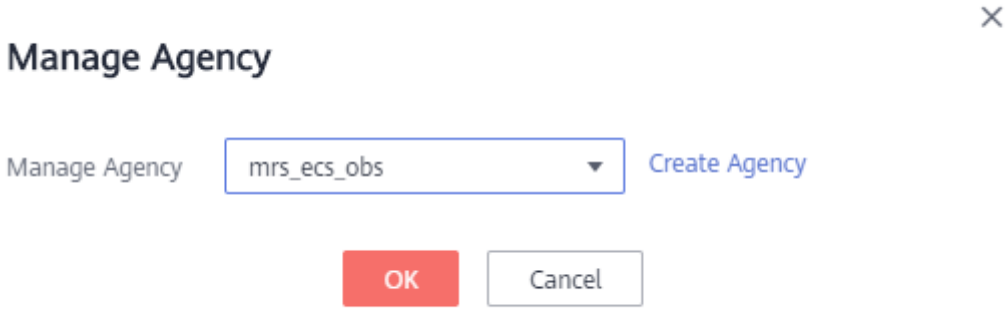
You can configure an agency when creating a cluster or bind an agency to an existing cluster to decouple storage and compute. This section uses an existing cluster as an example to describe how to configure an agency.

1. Log in to the MRS console. In the navigation pane on the left, choose **Clusters** > **Active Clusters**.

 Copyright © Huawei Technologies Co., Ltd.

2. Click the name of a cluster to go to the cluster details page.
3. On the **Dashboard** page, click **Synchronize** on the right side of **IAM User Sync** to synchronize IAM users.
4. On the **Dashboard** page, click **Manage Agency** on the right side of **Agency** to select the agency created in **Creating an ECS Agency**, and click **OK** to bind it to the cluster. Alternatively, click **Create Agency** to go to the IAM console to create an agency and bind it to the cluster.

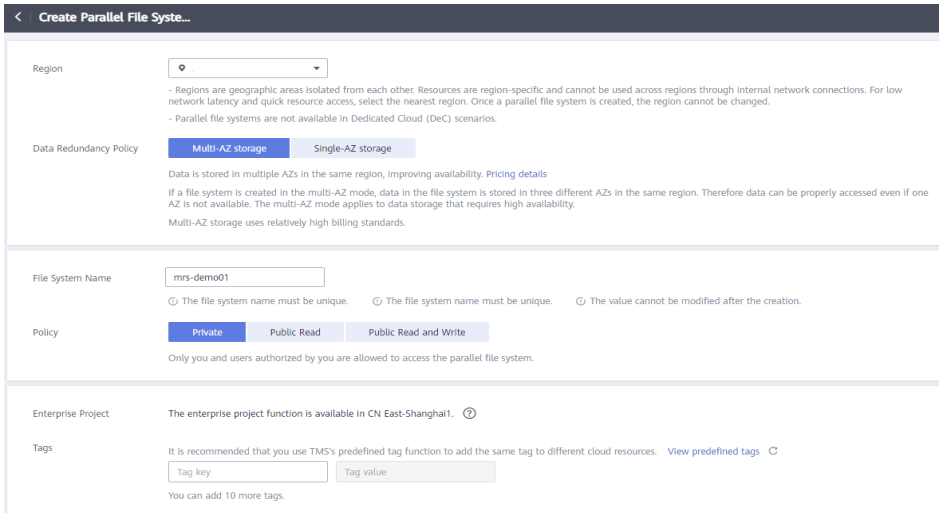
Figure 9-3 Binding an agency



Creating an OBS File System

1. Log in to the OBS console.
2. Choose **Parallel File System** > **Create Parallel File System**.
3. Enter the file system name, for example, **mrs-demo01**.

   Set other parameters as required.



4. Click **Create Now**.
5. In the parallel file system list on the OBS console, click a file system name to go to the details page.
6. In the navigation pane, choose **Files** and create **program** and **input** folders.
   - **program**: Upload the program package to this folder.

 Copyright © Huawei Technologies Co., Ltd.

- **input**: Upload the input data to this folder.

Accessing the OBS File System Through Hive

1. Log in to a master node as user **root**. For details, see **Logging In to an ECS**.
2. Verify that Hive can access OBS.
   a. Log in to the master node of the cluster as user **root** and run the following commands:

      cd /opt/Bigdata/client

      source bigdata_env

      source Hive/component_env

   b. View the list of files in file system **mrs-demo01**.

      hadoop fs -ls obs://mrs-demo01/

   c. Check whether the file list is returned. If it is returned, access to OBS is successful.

```
Found 2 items
drwxrwxrwx   - hive hive          0 2021-10-22 10:08 obs://mrs-demo01/input
drwxrwxrwx   - hive hive          0 2021-10-22 10:08 obs://mrs-demo01/program
```

   d. Run the following command to authenticate the user (skip this step for a normal cluster, that is, with Kerberos authentication disabled):

      kinit hive

      Enter the password of user **hive**. The default password is **Hive@123**. Change the password upon the first login.

   e. Run the Hive client command.

      beeline

   f. Access the OBS directory in the Beeline. For example, run the following command to create a Hive table and specify that data is stored in the **test_demo01** table of file system **mrs-demo01**:

      create table test_demo01(name string) location "obs://mrs-demo01/test_demo01";

   g. Run the following command to query all tables. If table **test_demo01** is displayed in the command output, access to OBS is successful.

      show tables;

```
+---------------+
|   tab_name    |
+---------------+
| test_demo01   |
+---------------+
1 row selected (0.301 seconds)
```

   h. Run the following command to check the table location.

      show create table test_demo01;

      Check whether the location of the table starts with **obs://***OBS bucket name***/**.

 Copyright © Huawei Technologies Co., Ltd.

```
  serialization.format' = ',' )
STORED AS INPUTFORMAT
  'org.apache.hadoop.mapred.TextInputFormat'
OUTPUTFORMAT
  'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTe
LOCATION
  'obs://mrs-demo01/test_demo01'
TBLPROPERTIES (
  'bucketing_version'='2',
  'transient_lastDdlTime'='1634872329')
```

i. Run the following command to write data into the table.

insert into test_demo01 values('mm'),('ww'),('ww');

Run the **select * from test_demo01;** command to check whether the data is written successfully.

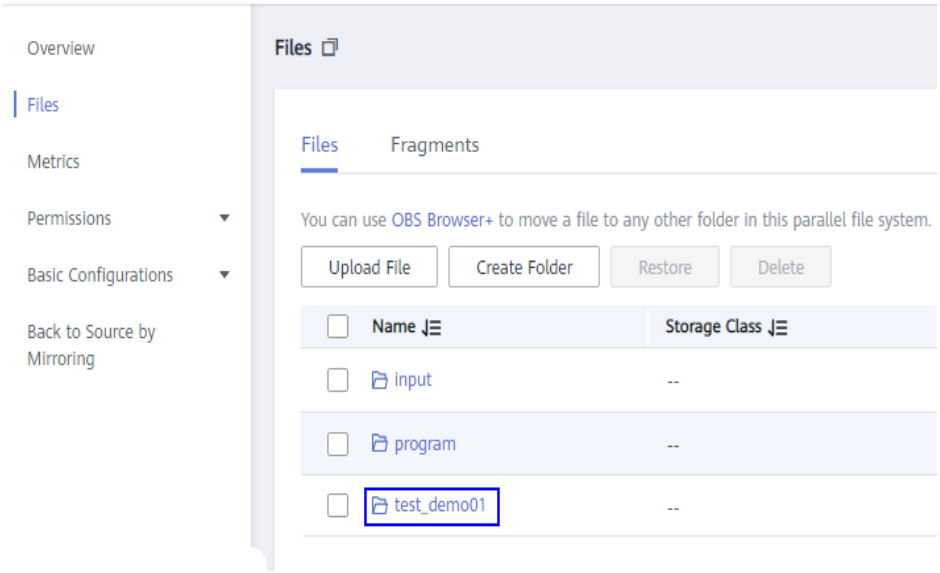
```
+-------------------+
| test_demo01.name  |
+-------------------+
| mm                |
| ww                |
| ww                |
+-------------------+
```

j. Run the **!q** command to exit the Beeline client.

k. Log in to the OBS console again.

l. Click **Parallel File System** and select the created file system.

m. Click **Files** to check whether the data exists in the created table.



 Copyright © Huawei Technologies Co., Ltd.

10 Submitting Spark Tasks to New Task Nodes

Add task nodes to a custom MRS cluster to increase compute capability. Task nodes are mainly used to process data instead of permanently storing data.

NOTE

Currently, task nodes can only be added to custom MRS clusters.

This section describes how to bind a new task node using tenant resources and submit Spark tasks to the new task node. You can get started by reading the following topics:

1. **Adding Task Nodes**
2. **Creating a Resource Pool**
3. **Creating a Tenant**
4. **Configuring Queues**
5. **Configuring Resource Distribution Policies**
6. **Creating a User**
7. **Using spark-submit to Submit a Task**
8. **Deleting Task Nodes**

Adding Task Nodes

1. On the details page of a custom MRS cluster, click the **Nodes** tab. On this tab page, click **Add Node Group**.
2. On the **Add Node Group** page that is displayed, set parameters as needed.

Table 10-1 Parameters for adding a node group

Parameter	Description
Instance Specifications	Select the flavor type of the hosts in the node group.

 Copyright © Huawei Technologies Co., Ltd.

Parameter	Description
Nodes	Configure the number of nodes in the node group.
System Disk	Configure the specifications and capacity of the system disks on the new nodes.
Data Disk (GB)/Disks	Set the specifications, capacity, and number of data disks of the new nodes.
Deploy Roles	Select **NM** to add a NodeManager role.

3. Click **OK**.

Creating a Resource Pool

Step 1 On the cluster details page, click **Tenants**.

Step 2 Click **Resource Pools**.

Step 3 Click **Create Resource Pool**.

Step 4 On the **Create Resource Pool** page, set the properties of the resource pool.

- **Name**: Enter the name of the resource pool, for example, **test1**.
- **Resource Label**: Enter the resource pool label, for example, **1**.
- **Available Hosts**: Enter the node added in **Adding Task Nodes**.

Step 5 Click **OK**.

----End

Creating a Tenant

Step 1 On the cluster details page, click **Tenants**.

Step 2 Click **Create Tenant**. On the displayed page, configure tenant properties.

Table 10-2 Tenant parameters

Parameter	Description
Name	Set the tenant name, for example, **tenant_spark**.
Tenant Type	Select **Leaf**. If **Leaf** is selected, the current tenant is a leaf tenant and no sub-tenant can be added. If **Non-leaf** is selected, sub-tenants can be added to the current tenant.
Dynamic Resource	If **Yarn** is selected, the system automatically creates a task queue using the tenant name in Yarn. If **Yarn** is not selected, the system does not automatically create a task queue.

 Copyright © Huawei Technologies Co., Ltd.

Parameter	Description
Default Resource Pool Capacity (%)	Set the percentage of computing resources used by the current tenant in the **default** resource pool, for example, **20%**.
Default Resource Pool Max. Capacity (%)	Set the maximum percentage of computing resources used by the current tenant in the **default** resource pool, for example, **80%**.
Storage Resource	If HDFS is selected, the system automatically creates the **/tenant** directory under the root directory of the HDFS when a tenant is created for the first time. If **HDFS** is not selected, the system does not create a storage directory under the root directory of the HDFS.
Maximum Number of Files/Directories	Set the maximum number of files or directories, for example, **100000000000**.
Storage Space Quota (MB)	Set the quota for using the storage space, for example, 50000 MB. This parameter indicates the maximum HDFS storage space that can be used by a tenant, but not the actual space used. If its value is greater than the size of the HDFS physical disk, the maximum space available is the full space of the HDFS physical disk. **NOTE** To ensure data reliability, the system automatically generates one backup file when a file is stored in the HDFS. That is, two replicas of the same file are stored by default. The HDFS storage space indicates the total disk space occupied by all these replicas. For example, if the value of **Storage Space Quota** is set to **500**, the actual space for storing files is about 250 MB (500/2 = 250).
Storage Path	Set the storage path, for example, **tenant/spark_test**. The system automatically creates a folder named after the tenant under the **/tenant** directory by default, for example, **spark_test**. The default HDFS storage directory for tenant **spark_test** is **tenant/spark_test**. When a tenant is created for the first time, the system creates the **/tenant** directory in the HDFS root directory. The storage path is customizable.
Services	Set other service resources associated with the current tenant. HBase is supported. To configure this parameter, click **Associate Services**. In the displayed dialog box, set **Service** to **HBase**. If **Association Mode** is set to **Exclusive**, service resources are occupied exclusively. If **share** is selected, service resources are shared.
Description	Enter the description of the current tenant.

 Copyright © Huawei Technologies Co., Ltd.

Step 3 Click **OK** to save the settings.

It takes a few minutes to save the settings. If the **Tenant created successfully** is displayed in the upper-right corner, the tenant is added successfully.

NOTE

- Roles, computing resources, and storage resources are automatically created when tenants are created.
- The new role has permissions on the computing and storage resources. The role and its permissions are controlled by the system automatically and cannot be controlled manually under **Manage Role**.
- If you want to use the tenant, create a system user and assign the Manager_tenant role and the role corresponding to the tenant to the user.

----End

Configuring Queues

Step 1 On the cluster details page, click **Tenants**.

Step 2 Click the **Queue Configuration** tab.

Step 3 In the tenant queue table, click **Modify** in the **Operation** column of the specified tenant queue.

NOTE

- In the tenant list on the left of the **Tenant Management** page, click the target tenant.

  In the displayed window, choose **Resource**. On the displayed page, click  to open the queue modification page.
- A queue can be bound to only one non-default resource pool.

By default, the resource tag is the one specified in **Creating a Resource Pool**. Set other parameters based on the site requirements.

Step 4 Click **OK**.

----End

Configuring Resource Distribution Policies

Step 1 On the cluster details page, click **Tenants**.

Step 2 Click **Resource Distribution Policies** and select the resource pool created in **Creating a Resource Pool**.

Step 3 Locate the row that contains **tenant_spark**, and click **Modify** in the **Operation** column.

- **Weight**: **20**
- **Minimum Resource**: **20**
- **Maximum Resource**: **80**
- **Reserved Resource**: **10**

 Copyright © Huawei Technologies Co., Ltd.

Step 4 Click **OK**.

----End

Creating a User

Step 1 Log in to FusionInsight Manager. For details, see **Accessing FusionInsight Manager**.

Step 2 Choose **System** > **Permission** > **User**. On the displayed page, click **Create User**.

- **Username**: **spark_test**
- **User Type**: **Human-Machine**
- **User Group**: hadoop and hive
- **Primary Group**: **hadoop**
- **Role**: **tenant_spark**

Step 3 Click **OK** to add the user.

----End

Using spark-submit to Submit a Task

1. Log in to the client node as user **root** and run the following commands:

   cd *Client installation directory*

   source bigdata_env

   source Spark2x/component_env

   For a cluster with Kerberos authentication enabled, run the **kinit spark_test** command. For a cluster with Kerberos authentication disabled, skip this step.

   Enter the password for authentication. Change the password upon the first login.

   cd Spark2x/spark/bin

   sh spark-submit --queue tenant_spark --class org.apache.spark.examples.SparkPi --master yarn-client ../examples/jars/spark-examples_*.jar

Deleting Task Nodes

1. On the cluster details page, click **Nodes**.
2. Locate the row that contains the target task node group, and click **Scale In** in the **Operation** column.
3. Set the **Scale-In Type** to **Specific node** and select the target nodes.

   NOTE

   The target nodes need to be shut down.

4. Select **I understand the consequences of performing the scale-in operation**, and click **OK**.

 Copyright © Huawei Technologies Co., Ltd.