MapReduce Service

Getting Started

 Issue
 01

 Date
 2025-01-23





HUAWEI CLOUD COMPUTING TECHNOLOGIES CO., LTD.

Copyright © Huawei Cloud Computing Technologies Co., Ltd. 2025. 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 Cloud Computing Technologies Co., Ltd.

Trademarks and Permissions

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

Notice

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

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

Huawei Cloud Computing Technologies Co., Ltd.

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

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

Contents

1 Creating and Using a Hadoop Cluster for Offline Analysis	1
2 Creating and Using a Kafka Cluster for Stream Processing	9
3 Creating and Using an HBase Cluster for Offline Query	18
4 Creating and Using a ClickHouse Cluster for Columnar Store	25
5 Creating and Using an MRS Cluster Requiring Security Authentication	32
6 Best Practices for Beginners	42

1 Creating and Using a Hadoop Cluster for Offline Analysis

Scenario

This topic describes how to create a Hadoop cluster for offline analysis and how to submit a wordcount job through the cluster client. A wordcount job is a classic Hadoop job that counts words in massive amounts of text.

The Hadoop cluster uses the open-source Hadoop ecosystem components, including YARN for cluster resource management, and Hive and Spark for offline large-scale distributed data storage and compute to provide massive data analysis and query capabilities.

Procedure

Before you start, complete operations described in **Preparations**. Then, follow these steps:

- 1. Creating an MRS cluster: Create a Hadoop analysis cluster of MRS 3.1.5.
- 2. Installing the Cluster Client: Download and install the MRS cluster client.
- 3. **Preparing Applications and Data**: Prepare the data files required for running the wordcount sample program on the MRS cluster client.
- 4. **Submitting a Job and Viewing the Result**: Submit a wordcount data analysis job on the cluster client and view the execution result.
- 5. **Releasing resources**: To avoid additional expenditures, release resources promptly if you no longer need them.

Preparations

• You have prepared an IAM user who has the permission to create MRS clusters. For details, see **Creating an MRS User**.

Step 1: Creating an MRS Cluster

Step 1 Go to the **Buy Cluster** page.

Step 2 Search for MapReduce Service in the service list and enter the MRS console.

Step 3 Click Buy Cluster. The **Quick Config** tab is displayed.

Step 4 Configure the cluster as you need. In this example, a pay-per-use MRS 3.1.5 cluster will be created. For more details about how to configure the parameters, see **Quickly Creating a Cluster**.

Parameter	Example Value	Description
Billing Mode	Pay-per-use	Billing mode of the cluster you want to create. MRS provides two billing modes: yearly/monthly and pay-per-use.
		Pay-per-use is a postpaid billing mode. You pay as you go and pay for what you use. The cluster usage is calculated by the second but billed every hour.
Region	CN-Hong Kong	Region where the MRS resources to be requested belong. MRS clusters in different regions cannot communicate with each other over an intranet. For lower network latency and quick resource access,
Cluster Name	mrs_demo	select the nearest region. Name of the MRS cluster you want to create.
Cluster Type	Custom	A range of clusters that accommodate diverse big data demands. You can select a Custom cluster to run a wide range of analytics components supported by MRS.
Version Type	Normal	Service type of the MRS
Cluster Version	MRS 3.1.5	Version of the MRS cluster. Supported open-source components and their functions vary depending on the cluster version. You are advised to select the latest version.
Component	Hadoop Analysis Cluster	Cluster templates containing preset opensource components you will need for your business.
AZ	AZ 1	Available AZ associated with the cluster region.
VPC	vpc-default	VPC where you want to create the cluster. You can click View VPC to view the name and ID. If no VPC is available, create one.

Table 1-1 MRS cluster parameters	
----------------------------------	--

Parameter	Example Value	Description
Subnet	subnet- default	Subnet where your cluster belongs. You can access the VPC management console to view the names and IDs of existing subnets in the VPC. If no subnet is created under the VPC, click Create Subnet to create one.
Cluster Node	Default value	Cluster node details.
Kerberos Authenticat ion	Disabled	Whether Kerberos authentication is enabled.
Username	admin/root	Username for logging in to the cluster management page and the ECS node.
Password	-	User password for logging in to the cluster management page and the ECS node.
Confirm Password	-	Enter the user password again.
Enterprise Project	default	Enterprise project to which the cluster belongs.
Secure Communic ations	Selected	Select the check box to agree to use the access control rules.

Figure 1-1 Buying a Hadoop analysis cluster

Cluster Configuration			
Cluster Name			
mrs			
The name must be unique. It can contain only 1 to 64 characters. Only letters, digits,	hyphens (-), and underscores (_) are allowed.		
Cluster Type			
Custom This has provide a such angue of customicable component combox. Component tables boccarge on different big data demande help you buy clusters faster.	Analysis cluster Subtate for analysis of processing maskine amounts of offline data to other much data offline processing basis usually regime large scale compute and stationary innorms. Data analysis components, such as Hadoop, Spark, Hölse, Hee, Flink, Duck and Tac are required.	Streaming cluster I Main the racic analysis of real-time data sources. I chreaming data and sources and the real-time data sources. Components such as Rafits and Flume are recommended.	Hybrid cluster Suitable for the filter data analysis and shrean processing. You can select analysis components such as Hadroon, Space, Haar, and Hive, we shream processing components such as Kafa and Flume.
Version Type () It Montal Normal Integrate MRS: makes and stable leatures and functions with open-source Clearly Version MRS 3.1.5 v Composed	e capabilities, offering high performance and stability		
Real-time Analysis Cluster Massive data collection, real-time data analysis and query Hadoop 3.1.1 Fink 1.1.2 Katha 2.11-2.4.0 ZooKeeper 3.6.3 Ranger 2.0.0 DickHouse 21.3.4.25	ClickHouse Cluster A Colum Database Management System (DBMS) for On-Line An ZooKeeper 3.6.3 ClickHouse 21.3.4.25	Hadoop Analysis Cluster Analysis and query of vata amounts of data Hadoop 3.11 Hive 3.10 Spark2.3.11 Tec 0.9.2 Fink 1.122 Zookkeeper 3.63 Ranger 2.00 Presto 333	HB3se Query Cluster Massive data storage and millisecond-level data quaries Hadoop 3.1.1 HBase 22.3 ZooKeeper 3.6.3 Ranger 2.0.0

- **Step 5** Click Buy Now. A page is displayed showing that the task has been submitted.
- **Step 6** Click **Back to Cluster List**. You can view the status of the newly created cluster on the **Active Clusters** page.

Wait for the cluster creation to complete. The initial status of the cluster is **Starting**. After the cluster is created, the cluster status becomes **Running**.

----End

Step 2: Installing the Cluster Client

You need to install a cluster client to connect to component services in the cluster, remotely access the client shell, and submit jobs.

The client can be installed on a node in or outside the cluster. This guide describes how to install the client on the Master1 node in the cluster.

- **Step 1** Click the MRS cluster name in the cluster list to go to the dashboard page.
- **Step 2** Click **Access Manager** next to **MRS Manager**. In the displayed dialog box, select **EIP** and configure the EIP information.

For the first access, click **Manage EIPs** to purchase an EIP on the EIP console. Go back to the **Access MRS Manager** dialog box, refresh the EIP list, and select the EIP.

Step 3 Select the confirmation check box and click **OK** to log in to the FusionInsight Manager of the cluster.

The username for logging in to FusionInsight Manager is **admin**, and the password is the one configured during cluster purchase.

Step 4 On the displayed **Homepage** page, click **••••** next to the cluster name and click **Download Client** to download the cluster client.

Figure 1-2 Downloading a client

Cluster

mrs_demo01 MRS	≈
Start	
Stop	
Restart	
Rolling-restart Service	
Synchronize Configurations	
Restart Configuration-Expired Instance	es
Health Check	
Download Client	

In the **Download Cluster Client** dialog box, set the following parameters:

- Set Select Client Type to Complete Client.
- For **Select Platform Type**, select the architecture of the node where the client is to be installed, for example, **x86_64**.

To check the architecture of a node in the cluster, click **Hosts** on FusionInsight Manager navigation pane on the top and click the target node name to go to the basic information page.

 Retain the default path for Save to Path. The generated file will be saved in the /tmp/FusionInsight-Client directory on the active OMS node (usually the Master1 node) of the cluster.

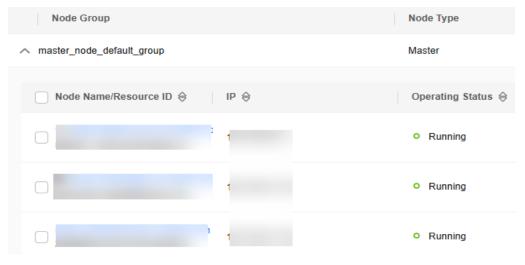
Figure 1-3 Downloading the cluster client

Download Cluster Client						
Download the mrs_demo01 client. The cluster client provides all services.						
Select Client Type:	Complete Client	Configuration Files Only				
Select Platform Type:	• x86_64 ○ aarch	64				
Save to Path:	/opt/Bigdata/client	0				
OK Cancel						

Click **OK** and wait until the client software is generated.

Step 5 Go back to the MRS console and click the cluster name in the cluster list. Go to the Nodes tab, click the name of the node that contains master1. In the upper right corner of the ECS details page, click Remote Login to log in to the Master1 node.

Figure 1-4 Checking the Master1 node



- **Step 6** Log in to the **Master1** node as user **root**. The password is the one you set for the **root** user during cluster purchase.
- **Step 7** Switch to the directory where the client software package is stored and decompress the package.

cd /tmp/FusionInsight-Client/

tar -xvf FusionInsight_Cluster_1_Services_Client.tar

tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar

Step 8 Go to the directory where the installation package is stored and install the client.

cd FusionInsight_Cluster_1_Services_ClientConfig

Install the client to a specified directory. (If the directory exists, it must be empty.)

For example, if the client is installed in the **/opt/client** directory, run the following command:

./install.sh /opt/client

Wait until the client installation is complete.

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

----End

Step 3: Preparing Applications and Data

You can run the wordcount sample program preset in the cluster client on the created cluster, or develop a big data application and upload it to the cluster.

This topic uses the wordcount sample program on the MRS cluster client as an example. You need to prepare the data files required for running the wordcount sample program.

- Step 1 Log in to the Master1 node as user root.
- Step 2 Prepare data files.

For example, the file names are **wordcount1.txt** and **wordcount2.txt**, and the content is as follows:

vi /opt/wordcount1.txt

hello word hello wordcount

vi /opt/wordcount2.txt

hello mapreduce hello hadoop

Step 3 Switch to the client installation directory, configure environment variables, and create an HDFS directory for storing sample data, for example, /user/example/ input.

cd /opt/client

source bigdata_env

hdfs dfs -mkdir /user/example/input

Step 4 Upload the sample data to HDFS.

hdfs dfs -put /opt/wordcount1.txt /user/example/input

hdfs dfs -put /opt/wordcount2.txt /user/example/input

----End

Step 4: Submitting a Job and Viewing the Result

- **Step 1** Log in to the client node (Master1) as user root.
- **Step 2** Submit a wordcount job, read source data for analysis, and output the execution result to the HDFS.

cd /opt/client

source bigdata_env

hadoop jar HDFS/hadoop/share/hadoop/mapreduce/hadoop-mapreduceexamples-3.3.1-*.jar wordcount "/user/example/input/*" "/user/example/ output/"

```
File Input Format Counters
Bytes Read=56
File Output Format Counters
Bytes Written=48
```

NOTE

- /user/example/output/ indicates the address for storing job output files on the HDFS. Set it to a directory that does not exist.
- The name of the **hadoop-mapreduce-examples-3.3.1-*.jar** file varies depending on the cluster client version. Use the actual name.

Step 3 Query job execution results.

1. Run the following command to view the job output file:

hdfs dfs -ls /user/example/output/

... /user/example/output/_SUCCESS ... /user/example/output/part-r-0000

2. The output is saved in the HDFS file system. You can run a command to download the output to the local PC and view it.

The following command is an example:

hdfs dfs -get /user/example/output/part-r-00000 /opt

cat /opt/part-r-00000

The content of the **part-r-00000** file is as follows:

hadoop 1 hello 4 mapreduce 1 word 1 wordcount 1 Step 4 View job run logs.

- 1. Log in to FusionInsight Manager of the target cluster as user **admin** and choose **Cluster** > **Services** > **Yarn**.
- 2. Click the **ResourceManager(xxx,Active)** link in the row where the **ResourceManager Web UI** is.
- 3. On the **All Applications** page, click the ID of the target job to view the job details.

On the **All Applications** page, you can confirm a task based on the task submission time and the user name that submits the task.

Figure 1-5 Checking job details

- Cluster	Cluster Metrics																
About	Apps Submitted	Apps Pending	Apps Running	,	Apps Complete	d	Contain	ers Running		Use	d Resources			1	otal Resour	ces	
Nodes Node Labels	6 0		0	6		0			<me< td=""><td>mory:0 B, vC</td><td>ores:0></td><td></td><td><memo< td=""><td>ry:144 GB,</td><td>vCores:96></td><td></td><td></td></memo<></td></me<>	mory:0 B, vC	ores:0>		<memo< td=""><td>ry:144 GB,</td><td>vCores:96></td><td></td><td></td></memo<>	ry:144 GB,	vCores:96>		
Applications	Cluster Nodes Metrics																
NEW NEW SAVING	Active Nodes	Decon	missioning Nodes		D	ecommission	ned Nodes		Lost	Nodes		Unhealth	y Nodes		Re	booted Nod	des
SUBMITTED	3 0				2				0		0			Q			
ACCEPTED RUNNING	User Metrics for mapred																
FINISHED	Apps Submitted Apps Pen	ding Apps Runn	ing Apps Comple	ted Co	tainers Running	Cont	ainers Pendin	g Cor	ntainers Reserv	ed Mei	mory Used	Memo	ry Pending	Memo	ory Reserve	d VCon	ies U
FAILED	0 0	0	0	0		0		0		0 B		0 B		0 B		0	
Scheduler	Scheduler Metrics																
	Scheduler Type	Scheduling Resou	rce Type	Minimum A	location		Ma	dmum Alloc	ation		Maxir	num Cluste	r Application	Priority	RM Dispa	tcher EventC	Queu
→ Tools	SuperiorYamScheduler [yam.io,	/gpu, memory-mb (u	init=Mi), vcores] <	memory:512,	vCores:1> <n< td=""><td>nemory:655</td><td>36, vCores:32,</td><td>yarn.io/gpu</td><td>: 9223372036</td><td>854775807></td><td>5</td><td></td><td></td><td></td><td>0</td><td></td><td></td></n<>	nemory:655	36, vCores:32,	yarn.io/gpu	: 9223372036	854775807>	5				0		
	Show 20 v entries																
	ID -	User QueueUser	Type	\$ Queue	° Priority ≎	0	FinishTime ¢	State °	FinalStatus ©		Allocated CPU VCores ©	Memory MB 0	Allocated GPUs ≎	Reserved CPU VCores	Memory MB 0	GPUs ©	
	application_1716278489865_0008	i root root	word MAPREE count	IUCE default	0	Tue May 21 16:53:09 +0800 2024	Tue May 21 16:53:23 +0800 2024	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.0

----End

Follow-up Operations: Releasing Resources

To avoid additional expenditures, release resources promptly if you no longer need them. For details, see **Deleting an MRS Cluster**.

Related Information

Hadoop components include HDFS, YARN, and MapReduce. You can run jobs to analyze or view offline data. For details, see **Using HDFS**, **Using MapReduce**, or **Using YARN**.

2 Creating and Using a Kafka Cluster for Stream Processing

Scenario

This topic helps you create a stream analysis cluster from scratch and generate and consume messages in a Kafka topic.

A Kafka cluster provides a message system with high throughput and scalability. It is widely used for log collection and monitoring data aggregation. Kafka is efficient in streaming data ingestion and real-time data processing and storage.

Procedure

Before you start, complete operations described in **Preparations**. Then, follow these steps:

- 1. **Creating an MRS Cluster**: Create a real-time analysis cluster of MRS 3.2.0-LTS.1.
- 2. Installing the Cluster Client: Download and install the MRS cluster client.
- 3. Using the Kafka Client to Create a Topic: Create a topic on the Kafka client.
- 4. **Managing Messages in a Kafka Topic**: Consume messages in a created topic on the Kafka client.
- 5. **Releasing resources**: To avoid additional expenditures, release resources promptly if you no longer need them.

Preparations

• You have prepared an IAM user who has the permission to create MRS clusters. For details, see **Creating an MRS User**.

Step 1: Creating an MRS Cluster

- Step 1 Go to the Buy Cluster page.
- **Step 2** Search for MapReduce Service in the service list and enter the MRS console.
- Step 3 Click Buy Cluster. The Quick Config tab is displayed.

Step 4 Configure the cluster as you need. In this example, a pay-per-use MRS 3.2.0-LTS.1 cluster will be created. For more details about how to configure the parameters, see **Quickly Creating a Cluster**.

Parameter	Example Value	Description
Billing Mode	Pay-per-use	Billing mode of the cluster you want to create. MRS provides two billing modes: yearly/monthly and pay-per-use. Pay-per-use is a postpaid billing mode. You pay as you go and pay for what you use. The cluster usage is calculated by the second but billed every hour.
Region	CN-Hong Kong	Region where the MRS resources to be requested belong. MRS clusters in different regions cannot communicate with each other over an intranet. For lower network latency and quick resource access, select the nearest region.
Cluster Name	mrs_demo	Name of the MRS cluster you want to create.
Cluster Type	Custom	A range of clusters that accommodate diverse big data demands. You can select a Custom cluster to run a wide range of analytics components supported by MRS.
Version Type	LTS	Version of the MRS cluster. Supported open-source components and their functions vary depending on the cluster version. You are advised to select the latest version.
Cluster Version	MRS 3.2.0- LTS.1	Service type of the MRS
Component	Real-time Analysis Cluster	Cluster templates containing preset opensource components you will need for your business.
AZ	AZ 1	Available AZ associated with the cluster region.
VPC	vpc-default	VPC where you want to create the cluster. You can click View VPC to view the name and ID. If no VPC is available, create one.
Subnet	subnet- default	Subnet where your cluster belongs. You can access the VPC management console to view the names and IDs of existing subnets in the VPC. If no subnet is created under the VPC, click Create Subnet to create one.

Parameter	Example Value	Description
Cluster Node	Default value	Cluster node details.
Kerberos Authenticat ion	Disabled	Whether Kerberos authentication is enabled.
Username	admin/root	Username for logging in to the cluster management page and the ECS node.
Password	-	User password for logging in to the cluster management page and the ECS node.
Confirm Password	-	Enter the user password again.
Enterprise Project	default	Enterprise project to which the cluster belongs.
Secure Communic ations	Selected	Select the check box to agree to use the access control rules.

Figure 2-1 Purchasing a real-time analysis cluster



- **Step 5** Click Buy Now. A page is displayed showing that the task has been submitted.
- **Step 6** Click **Back to Cluster List**. You can view the status of the newly created cluster on the **Active Clusters** page.

Wait for the cluster creation to complete. The initial status of the cluster is **Starting**. After the cluster is created, the cluster status becomes **Running**.

----End

Step 2: Installing the Cluster Client

You need to install a cluster client to connect to component services in the cluster and submit jobs.

You can install the client on a node in or outside the cluster. This topic installs the client on the **Master1** node as an example.

- **Step 1** Click the MRS cluster name in the cluster list to go to the dashboard page.
- **Step 2** Click **Access Manager** next to **MRS Manager**. In the displayed dialog box, select **EIP** and configure the EIP information.

For the first access, click **Manage EIPs** to purchase an EIP on the EIP console. Go back to the **Access MRS Manager** dialog box, refresh the EIP list, and select the EIP.

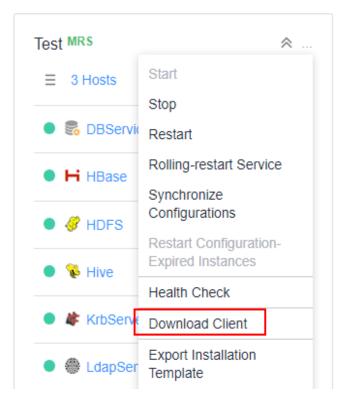
Step 3 Select the confirmation check box and click **OK** to log in to the FusionInsight Manager of the cluster.

The username for logging in to FusionInsight Manager is **admin**, and the password is the one configured during cluster purchase.

Step 4 On the displayed **Homepage** page, click •••• next to the cluster name and click **Download Client** to download the cluster client.

Figure 2-2 Downloading the client

Cluster



In the **Download Cluster Client** dialog box, set the following parameters:

- Set Select Client Type to Complete Client.
- Retain the default value for **Platform Type**, for example, **x86_64**.

• Retain the default path for **Save to Path**. The generated file will be saved in the **/tmp/FusionInsight-Client** directory on the active OMS node of the cluster.

Figure 2-3 Downloading the cluster client

Download Cluster Client

Download the client.	The cluster client provides	all services.
Select Client Type:	Complete Client	Configuration Files Only
Select Platform Type:	● x86_64	h64
Save to Path :	/tmp/FusionInsight-Client/	0
	ОК	Cancel

Click **OK** and wait until the client software is generated.

Step 5 Go back to the MRS console and click the cluster name in the cluster list. Go to the Nodes tab, click the name of the node that contains master1. In the upper right corner of the ECS details page, click Remote Login to log in to the Master1 node.

Figure 2-4 Checking the Master1 node

Service ZooKeeper / Instances

Service Status Instances Service Config Add Instance More V	guration		
🗌 Role 🔶	Host Name	OM IP Address	Business IP Address
quorumpeer	node-master2ggzl.mrs-82mt.com	192.1	192. ⁻
quorumpeer	node-master3PywB.mrs-82mt.com	192.1	192.1
quorumpeer	node-master1cwUo.mrs-82mt.com	192.1	192.1

- **Step 6** Log in to the **Master1** node as user **root**. The password is the one you set for the **root** user during cluster purchase.
- **Step 7** Switch to the directory where the client software package is stored and decompress the package.

cd /tmp/FusionInsight-Client/

tar -xvf FusionInsight_Cluster_1_Services_Client.tar

tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar

Step 8 Go to the directory where the installation package is stored and install the client.

cd FusionInsight_Cluster_1_Services_ClientConfig

Install the client to a specified directory (an absolute path), for example, **/opt/ client**.

./install.sh /opt/client

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

NOTE

A client installation directory will be automatically created if it does not exist. If there is such directory, it must be empty. The directory name cannot contain spaces. The client installation directory can contain only uppercase letters, lowercase letters, digits, and underscores (_).

----End

Step 3: Using the Kafka Client to Create a Topic

- **Step 1** In the cluster list, click the name of the target cluster. The dashboard tab is displayed.
- **Step 2** On the displayed page, click **Synchronize** next to **IAM User Sync**. In the displayed dialog box, select **All**, and click **Synchronize**. Wait until the synchronization task is complete.
- **Step 3** Go to the **Components** tab, click **ZooKeeper**, and then click the **Instances** tab. Check and record the IP address of a ZooKeeper quorumpeer role instance.

Figure 2-5 Checking IP addresses of ZooKeeper role instances

Service ZooKeeper / Instances			
Service Status Instances Se	rvice Configuration		
Add Instance More	×		
🗌 Role 🔶	Host Name \ominus	OM IP Address	Business IP Address
quorumpeer	node-master2ggzl.mrs-82mt.com	192.1	192.
quorumpeer	node-master3PywB.mrs-82mt.com	192.1	192.
quorumpeer	node-master1cwUo.mrs-82mt.com	192.1	192. ⁻

- **Step 4** Click **Service Configuration** and check the value of **clientPort**, which indicates the ZooKeeper client connection port.
- **Step 5** Click **Service ZooKeeper** to return to the component list.

Figure 2-6 Going back to the component list

ervice Status	Instances	Service Configuration	

Step 6 Click **Kafka**, and then the **Instances** tab. Check and record the IP addresses of a Kafka Broker instance.

Figure 2-7 Checking the IP address of a broker instance

Service Kafka / Instances			
Service Status Instances	Service Configuration		
More v			
🗌 Role 🔶	Host Name	OM IP Address	Business IP Address
Broker	node-master2ggzl.mrs-82mt.com	192.	192.1
Broker	node-master3PywB.mrs-82mt.com	192.	192.1
Broker	test1ESxJ.mrs-82mt.com	192.	192.

- **Step 7** Click **Service Configuration** and check the value of **port**, which indicates the port for connecting to Kafka Broker.
- Step 8 Log in to the node (Master1) where the MRS client is located as user root.
- **Step 9** Switch to the client installation directory and configure environment variables.

cd /opt/client

source bigdata_env

Step 10 Create a Kafka topic.

kafka-topics.sh --create --zookeeper *IP* address of ZooKeeper role instance.ZooKeeper client connection port /kafka --partitions 2 --replicationfactor 2 --topic Topic name

The following is an example:

kafka-topics.sh --create --zookeeper 192.168.21.234:2181/kafka --partitions 2 --replication-factor 2 --topic Topic1

If the following information is displayed, the topic is created:

Created topic Topic1.

----End

Step 4: Managing Messages in the Kafka Topic

- **Step 1** Log in to the node (**Master1**) where the MRS client is deployed as user **root**.
- **Step 2** Switch to the client installation directory and configure environment variables.

cd /opt/client

source bigdata_env

Step 3 Generate a message in Topic1.

kafka-console-producer.sh --broker-list *IP address of the node where the Kafka Broker role is deployed*:*Broker connection port* --topic *Topic name* -producer.config /opt/hadoopclient/Kafka/kafka/config/producer.properties

For the IP address and port number of the node where the Kafka Broker instance is deployed, see **Step 6** and **Step 7** in **Step 3**: **Using the Kafka Client to Create a Topic**.

The following is an example:

kafka-console-producer.sh --broker-list 192.168.21.21:9092 --topic Topic1 -producer.config /opt/client/Kafka/kafka/config/producer.properties

Step 4 Open a new client connection window.

cd /opt/client

source bigdata_env

Step 5 Consume messages in Topic1.

kafka-console-consumer.sh --topic *Topic name* --bootstrap-server *IP address of the node where the Kafka Broker role is deployed*.*Broker connection port* -consumer.config /opt/client/Kafka/kafka/config/consumer.properties

The following is an example:

kafka-console-consumer.sh --topic Topic1 --bootstrap-server 192.168.21.21:9092 --consumer.config /opt/client/Kafka/kafka/config/ consumer.properties

Step 6 Enter some content in the command line that generates messages in **Step 3**. The content is used as the messages generated by the producer. Press **Enter** to send the message.

The following is an example:

>aaa >bbb >ccc

To stop generating messages, press **Ctrl+C** to exit.

- **Step 7** In the message consuming window of **Step 5**, check whether the messages are consumed.
 - aaa bbb ccc

----End

Follow-up Operations: Releasing Resources

To avoid additional expenditures, release resources promptly if you no longer need them. For details, see **Deleting an MRS Cluster**.

Related Information

For information about Kafka permission management, topic management and message consumption, HA configuration, and data balancing, see Using Kafka.

3 Creating and Using an HBase Cluster for Offline Query

Scenario

This topic helps you create an HBase query cluster from scratch and describes how to create and query HBase tables through the cluster client.

An HBase cluster uses Hadoop and HBase components to provide a columnoriented distributed cloud storage system featuring enhanced reliability, great performance, and elastic scalability. It applies to the storage and distributed computing of massive amounts of data. You can use HBase to build a storage system capable of storing TB- or even PB-level data. With HBase, you can filter and analyze data with ease and get responses in milliseconds, rapidly mining data value.

Procedure

Before you start, complete operations described in **Preparations**. Then, follow these steps:

- 1. **Creating an MRS Cluster**: Create an HBase query cluster of MRS 3.2.0-LTS.1.
- 2. Installing the Cluster Client: Download and install the MRS cluster client.
- 3. **Creating a Table Using the HBase Client**: Create a table, insert table data, query the data, and delete the table on the HBase client.
- 4. **Releasing resources**: To avoid additional expenditures, release resources promptly if you no longer need them.

Preparations

• You have prepared an IAM user who has the permission to create MRS clusters. For details, see **Creating an MRS User**.

Step 1: Creating an MRS Cluster

Step 1 Go to the **Buy Cluster** page.

Step 2 Search for MapReduce Service in the service list and enter the MRS console.

Step 3 Click Buy Cluster. The **Quick Config** tab is displayed.

Step 4 Configure the cluster as you need. In this example, a pay-per-use MRS 3.2.0-LTS.1 cluster will be created. For more details about how to configure the parameters, see **Quickly Creating a Cluster**.

Parameter	Example Value	Description
Billing Mode	Pay-per-use	Billing mode of the cluster you want to create. MRS provides two billing modes: yearly/monthly and pay-per-use. Pay-per-use is a postpaid billing mode. You pay as you go and pay for what you use. The cluster usage
		is calculated by the second but billed every hour.
Region	CN-Hong Kong	Region where the MRS resources to be requested belong.
		MRS clusters in different regions cannot communicate with each other over an intranet. For lower network latency and quick resource access, select the nearest region.
Cluster Name	mrs_demo	Name of the MRS cluster you want to create.
Cluster Type	Custom	A range of clusters that accommodate diverse big data demands. You can select a Custom cluster to run a wide range of analytics components supported by MRS.
Version Type	LTS	Version of the MRS cluster. Supported open-source components and their functions vary depending on the cluster version.
Cluster Version	MRS 3.2.0- LTS.1	Version of the MRS cluster. Supported open-source components and their functions vary depending on the cluster version. You are advised to select the latest version.
Component	HBase Query Cluster	Cluster templates containing preset opensource components you will need for your business.
AZ	AZ 1	Available AZ associated with the cluster region.
VPC	vpc-default	VPC where you want to create the cluster. You can click View VPC to view the name and ID. If no VPC is available, create one.

Parameter	Example Value	Description
Subnet	subnet- default	Subnet where your cluster belongs. You can access the VPC management console to view the names and IDs of existing subnets in the VPC. If no subnet is created under the VPC, click Create Subnet to create one.
Cluster Node	Default value	Cluster node details.
Kerberos Authenticat ion	Disabled	Whether Kerberos authentication is enabled.
Username	admin/root	Username for logging in to the cluster management page and the ECS node.
Password	-	User password for logging in to the cluster management page and the ECS node.
Confirm Password	-	Enter the user password again.
Enterprise Project	default	Enterprise project to which the cluster belongs.
Secure Communic ations	Selected	Select the check box to agree to use the access control rules.

Figure 3-1 Purchasing an HBase query cluster

Cluster Configuration			
Cluster Name			
ms_			
The name must be unique. It can contain only 1 to 64 characters. Only letters, digits,	hyphens (-), and underscores (_) are allowed.		
Cluster Type			
Ceston This type provides a welde tange of customicable component combox. Component memory on different big data demands help you buy clusters tester.	Analysis cluster Subtain to analyzing and processing massive amounts of offline data to ottain read data. Offline processing tables unable registrate company and strange resource. Oral analysis components, such as Haboo, Spark, HBase, Hier, Flork, Occioe and hie are regarded.	Streaming cluster I Ideal for guick analysis of read-time data sources. I charaming data processing usually hais High CPU and memory requirements. Components such as Kalita and Flume are recommended.	Hypoid cluster Subliab for chird filling data analysis and stream processing Visio can select adaptive components such as Hadroos, Space, Hellas, and Hive, and Stream processing components such as Katha and Faure.
Version Type O LTD Vormal Long Thes Boyce (LTS) employs MRS's own components to provide highly reliable Long Test Boyce (LTS) employs MRS's own components to provide highly reliable MRS 322-B-LTS.1 v Component	clusters with shong DR capabilities, making long-term support and evolution possib	a	
Real-time Analysis Cluster Matsive data collection, real-time data analysis and query Hadoup 3.11 First 1.15.0 Katha 2.11:2.4.0 ZooKeeper 3.6.3 Ranger 2.0.0 ClickHoure 22.3.22	ClickHouse Cluster A Column Database Management System (DBMS) for On-Line An Zookkeeper 3.6.3 ClickHouse 22.1.2.2	Hadoop Analysis Cluster Analysis and gavey of vatal amounts of data Hadoop 3.1.1 Hvrs 1.1.0 Spenk2 3.1.1 Fink 1.15.0 Zeekkeeper 3.5.3 Ranger 2.0.0 Tec 9.9.2	HBase Query Cluster Massive data storage and mittsecond-level data quaries Hedoop 1.1.1 HBase 2.2.3 Zookkeper 3.6.3 Ranger 2.0.0

- **Step 5** Click Buy Now. A page is displayed showing that the task has been submitted.
- **Step 6** Click **Back to Cluster List**. You can view the status of the newly created cluster on the **Active Clusters** page.

Wait for the cluster creation to complete. The initial status of the cluster is **Starting**. After the cluster is created, the cluster status becomes **Running**.

----End

Step 2: Installing the Cluster Client

You need to install a cluster client to connect to component services in the cluster and submit jobs.

You can install the client on a node in or outside the cluster. This topic installs the client on the **Master1** node as an example.

- **Step 1** Click the MRS cluster name in the cluster list to go to the dashboard page.
- **Step 2** Click **Access Manager** next to **MRS Manager**. In the displayed dialog box, select **EIP** and configure the EIP information.

For the first access, click **Manage EIPs** to purchase an EIP on the EIP console. Go back to the **Access MRS Manager** dialog box, refresh the EIP list, and select the EIP.

Step 3 Select the confirmation check box and click **OK** to log in to the FusionInsight Manager of the cluster.

The username for logging in to FusionInsight Manager is **admin**, and the password is the one configured during cluster purchase.

Step 4 On the displayed **Homepage** page, click •••• next to the cluster name and click **Download Client** to download the cluster client.

Test MRS \approx Start 3 Hosts = Stop 🖏 DBServi Restart Rolling-restart Service H HBase Synchronize Configurations 💡 HDFS Restart Configuration-Expired Instances 🚯 Hive Health Check KrbSer Download Client Export Installation LdapSer Template

Figure 3-2 Downloading the client

Cluster

In the **Download Cluster Client** dialog box, set the following parameters:

- Set Select Client Type to Complete Client.
- Retain the default value for **Platform Type**, for example, **x86_64**.
- Retain the default path for **Save to Path**. The generated file will be saved in the **/tmp/FusionInsight-Client** directory on the active OMS node of the cluster.

Figure 3-3 Downloading the cluster client

Download Cluster Client

Download the client.	The cluster client provides	all services.
Select Client Type:	Complete Client	Configuration Files Only
Select Platform Type:	● x86_64	:h64
Save to Path :	/tmp/FusionInsight-Client/	0
	ОК	Cancel

Click **OK** and wait until the client software is generated.

Step 5 Go back to the MRS console and click the cluster name in the cluster list. Go to the Nodes tab, click the name of the node that contains master1. In the upper right corner of the ECS details page, click Remote Login to log in to the Master1 node.

Figure 3-4 Checking the Master1 node

Service ZooKeeper / Instances				
Service Status Instances Serv	vice Configuration			
Add Instance More V				
Role	Host Name	OM IP Address	Business IP Address	
quorumpeer	node-master2ggzl.mrs-82mt.com	192.1	192. ⁻	
quorumpeer	node-master3PywB.mrs-82mt.com	192.1	192. ⁻	
quorumpeer	node-master1cwUo.mrs-82mt.com	192.1	192.1	

- **Step 6** Log in to the **Master1** node as user **root**. The password is the one you set for the **root** user during cluster purchase.
- **Step 7** Switch to the directory where the client software package is stored and decompress the package.

cd /tmp/FusionInsight-Client/

tar -xvf FusionInsight_Cluster_1_Services_Client.tar

tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar

Step 8 Go to the directory where the installation package is stored and install the client.

cd FusionInsight_Cluster_1_Services_ClientConfig

Install the client to a specified directory (an absolute path), for example, **/opt/ client**.

./install.sh /opt/client

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

NOTE

A client installation directory will be automatically created if it does not exist. If there is such directory, it must be empty. The directory name cannot contain spaces. The client installation directory can contain only uppercase letters, lowercase letters, digits, and underscores (_).

----End

Step 3: Creating a Table Using the HBase Client

- **Step 1** Log in to the node (Master1) where the MRS client is deployed as user root.
- **Step 2** Switch to the client installation directory and configure environment variables.

cd /opt/client

source bigdata_env

Step 3 Access the HBase shell CLI.

hbase shell

- **Step 4** Create the **user_info** table on the HBase client.
 - Create the user_info table. create 'user_info,{NAME => 'i'}
 - Add data to the user_info table.
 put 'user_info','12005000201','i:name','A'
 put 'user_info','12005000201','i:gender','Male'
 put 'user_info','12005000201','i:address','City A'
 put 'user_info','12005000201','i:ddress','City A'
 put 'user_info','12005000201','i:degree','master'
 put 'user_info','12005000201','i:pose','manager'
- **Step 5** Query the HBase table.
 - 1. Query usernames and addresses by user ID. scan 'user_infd', {STARTROW=>'12005000201',STOPROW=>'12005000201',COLUMNS=>['i:name','i:address']} The query result is as follows: 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	

 Query information by username. scan 'user_info',{FILTER=>"SingleColumnValueFilter('i','name',=,'binary:A')"}

The query result is as follows:

ROW +CELL	COLUMN
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	

Step 6 Delete the HBase table.

- 1. Delete user data from the user information table. delete 'user_info',' 12005000201','i'
- Delete the user information table. disable 'user_info' drop 'user_info'

----End

Follow-up Operations: Releasing Resources

To avoid additional expenditures, release resources promptly if you no longer need them. For details, see **Deleting an MRS Cluster**.

Related Information

For details about HBase permission management, indexes and global secondary indexes, and HBase data migration using BulkLoad, see **Using HBase**.

4 Creating and Using a ClickHouse Cluster for Columnar Store

Scenario

This topic helps you create a ClickHouse cluster from scratch and create and query a ClickHouse table through the cluster client.

ClickHouse is an open-source columnar database oriented to online analysis and processing. It is independent of the Hadoop big data system and features ultimate compression rate and fast query performance.

Procedure

Before you start, complete operations described in **Preparations**. Then, follow these steps:

- 1. **Creating an MRS Cluster**: Create a ClickHouse cluster of MRS 3.2.0-LTS.1.
- 2. Installing the Cluster Client: Download and install the MRS cluster client.
- 3. **Creating a Table through the ClickHouse Client**: Create a table on the ClickHouse client and insert data into the table.
- 4. **Releasing resources**: To avoid additional expenditures, release resources promptly if you no longer need them.

Preparations

• You have prepared an IAM user who has the permission to create MRS clusters. For details, see **Creating an MRS User**.

Step 1: Creating an MRS Cluster

- **Step 1** Go to the **Buy Cluster** page.
- **Step 2** Search for MapReduce Service in the service list and enter the MRS console.
- **Step 3** Click Buy Cluster. The **Quick Config** tab is displayed.
- **Step 4** Configure the cluster as you need. In this example, a pay-per-use MRS 3.2.0-LTS.1 cluster will be created. For more details about how to configure the parameters, see **Quickly Creating a Cluster**.

Parameter	Example Value	Description
Billing Mode	Pay-per-use	Billing mode of the cluster you want to create. MRS provides two billing modes: yearly/monthly and pay-per-use.
		Pay-per-use is a postpaid billing mode. You pay as you go and pay for what you use. The cluster usage is calculated by the second but billed every hour.
Region	CN-Hong Kong	Region where the MRS resources to be requested belong.
		MRS clusters in different regions cannot communicate with each other over an intranet. For lower network latency and quick resource access, select the nearest region.
Cluster Name	mrs_demo	Name of the MRS cluster you want to create.
Cluster Type	Custom	A range of clusters that accommodate diverse big data demands. You can select a Custom cluster to run a wide range of analytics components supported by MRS.
Version Type	LTS	Service type of the MRS
Cluster Version	MRS 3.2.0- LTS.1	Version of the MRS cluster. Supported open-source components and their functions vary depending on the cluster version. You are advised to select the latest version.
Component	ClickHouse Cluster	Cluster templates containing preset opensource components you will need for your business.
AZ	AZ 1	Available AZ associated with the cluster region.
VPC	vpc-default	VPC where you want to create the cluster. You can click View VPC to view the name and ID. If no VPC is available, create one.
Subnet	subnet- default	Subnet where your cluster belongs. You can access the VPC management console to view the names and IDs of existing subnets in the VPC. If no subnet is created under the VPC, click Create Subnet to create one.
Cluster Node	Default value	Cluster node details.
Kerberos Authenticat ion	Disabled	Whether Kerberos authentication is enabled.

Table 4-1	MRS	cluster	parameters
-----------	-----	---------	------------

Parameter	Example Value	Description
Username	admin/root	Username for logging in to the cluster management page and the ECS node.
Password	-	User password for logging in to the cluster management page and the ECS node.
Confirm Password	-	Enter the user password again.
Enterprise Project	default	Enterprise project to which the cluster belongs.
Secure Communic ations	Selected	Select the check box to agree to use the access control rules.

Figure 4-1 Buying a ClickHouse cluster

Cluster Configuration			
Cluster Name			
mrs			
The name must be unique. It can contain only 1 to 64 characters. Only letters, digit	s, hyphens (-), and underscores (_) are allowed.		
Cluster Type			
Centon • The top provide a wide range of cultomicable component controls. • The top provides a wide range of additional component top data demands help you buy cluster facility.	Analysis cluster educate for analyzing and processing massive amounts of offline data to drain results by a second	Streaming cluster	Hypod cluster Subalish for shaft analysis and stream processing. You can select adaptive components such as Hodikas, Spark Hiber, and Hive, and stream processing components such as Kofak and Flume.
Version Type ① Ity Tormal Ity Tormal Ity Tormal Ity Tormal Ity	e dusters with strong DR capabilities, making long-term support and evolution possib	x	
Real-time Analysis Cluster Mastive data collection, real-time data analysis and query Hadoop 13.1 Film: 1150 Katha 2.11:24.0 ZoalKaeper 16.3 Ranger 2.00 ClosHouse 22.32.2 ClosHouse 23.32 ClosHouse 23.32	ClickHouse Cluster A Colum Database Management System (DBMS) for On-Line An Zookeeper 1.6.1 ClickHouse 22.3.2.2	Hadoop Analysis Cluster Analysis and query di vata amounts of data Hadoop 13.1 Hive 1.10 Spanic2 13.11 Piek 115.0 Zookesper 3.63 Ranger 2.00 Tec 0.9.2	HBase Query Cluster Mastive data storage and millisecond-aveil data queries Hadoop 1.1.1 HBase 2.2.3 ZooKeeper 3.6.3 Ranger 2.0.0

- **Step 5** Click Buy Now. A page is displayed showing that the task has been submitted.
- **Step 6** Click **Back to Cluster List**. You can view the status of the newly created cluster on the **Active Clusters** page.

Wait for the cluster creation to complete. The initial status of the cluster is **Starting**. After the cluster is created, the cluster status becomes **Running**.

----End

Step 2: Installing the Cluster Client

You need to install a cluster client to connect to component services in the cluster and submit jobs.

You can install the client on a node in or outside the cluster. This topic installs the client on the **Master1** node as an example.

- **Step 1** Click the MRS cluster name in the cluster list to go to the dashboard page.
- **Step 2** Click **Access Manager** next to **MRS Manager**. In the displayed dialog box, select **EIP** and configure the EIP information.

For the first access, click **Manage EIPs** to purchase an EIP on the EIP console. Go back to the **Access MRS Manager** dialog box, refresh the EIP list, and select the EIP.

Step 3 Select the confirmation check box and click **OK** to log in to the FusionInsight Manager of the cluster.

The username for logging in to FusionInsight Manager is **admin**, and the password is the one configured during cluster purchase.

Step 4 On the displayed **Homepage** page, click •••• next to the cluster name and click **Download Client** to download the cluster client.

Figure 4-2 Downloading the client

Cluster

Test MRS	≈
≡ 3 Hosts	Start
	Stop
🔍 👼 DBServio	Restart
• H HBase	Rolling-restart Service
	Synchronize
🔍 🦪 HDFS	Configurations
🔵 😵 Hive	Restart Configuration- Expired Instances
	Health Check
🔍 🍂 KrbServe	Download Client
• 🏶 LdapSer	Export Installation Template

In the **Download Cluster Client** dialog box, set the following parameters:

- Set Select Client Type to Complete Client.
- Retain the default value for **Platform Type**, for example, **x86_64**.
- Retain the default path for **Save to Path**. The generated file will be saved in the **/tmp/FusionInsight-Client** directory on the active OMS node of the cluster.

0

Cancel

Figure 4-3 Downloading the cluster client

Download Cluster Client

Save to Path :

Download the clien	t. The cluster clie	ent provides	s all services.
Select Client Type:	Complete	Client	Configuration Files Only
Select Platform Type:	● x86_64	aarc	:h64

/tmp/FusionInsight-Client/

ΟK

Click **OK** and wait until the client software is generated.

Step 5 Go back to the MRS console and click the cluster name in the cluster list. Go to the Nodes tab, click the name of the node that contains master1. In the upper right corner of the ECS details page, click Remote Login to log in to the Master1 node.

Figure 4-4 Checking the Master1 node

Service ZooKeeper / Instances	Service ZooKeeper / Instances				
Service Status Instances Ser	vice Configuration				
Add Instance More					
🗌 Role 🔶	Host Name	OM IP Address	Business IP Address		
quorumpeer	node-master2ggzl.mrs-82mt.com	192.1	192. ⁻		
quorumpeer	node-master3PywB.mrs-82mt.com	192.1	192.1		
quorumpeer	node-master1cwUo.mrs-82mt.com	192.1	192.1		

- **Step 6** Log in to the **Master1** node as user **root**. The password is the one you set for the **root** user during cluster purchase.
- **Step 7** Switch to the directory where the client software package is stored and decompress the package.

cd /tmp/FusionInsight-Client/

tar -xvf FusionInsight_Cluster_1_Services_Client.tar

tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar

Step 8 Go to the directory where the installation package is stored and install the client.

cd FusionInsight_Cluster_1_Services_ClientConfig

Install the client to a specified directory (an absolute path), for example, **/opt/ client**.

./install.sh /opt/client

... component client is installed successfully

NOTE

A client installation directory will be automatically created if it does not exist. If there is such directory, it must be empty. The directory name cannot contain spaces. The client installation directory can contain only uppercase letters, lowercase letters, digits, and underscores (_).

----End

Step 3: Creating a Table Through the ClickHouse Client

- Step 1 Log in to the node (Master1) where the MRS client is deployed as user root.
- **Step 2** Switch to the client installation directory and configure environment variables.

cd /opt/client

source bigdata_env

Step 3 Run the clickhouse client command to connect to the ClickHouse server.

clickhouse client --host *IP* address of the ClickHouseServer instance --port 9000 --user Username --password

NOTE

- To obtain the IP address of the ClickHouseServer instance, log in to FusionInsight Manager of the cluster and choose **Cluster** > **Services** > **ClickHouse** > **Instances**.
- Clusters with Kerberos authentication disabled use non-SSL connections by default. The default connection port is 9000. To view the port number, log in to FusionInsight Manager of the cluster, choose Cluster > Services > ClickHouse > Configurations, and search for tcp_port.
- If the --user and --password parameters are not specified, the default user is used for logging in to the ClickHouse client by default. If you want to specify the username and password, run the create user SQL statement on the ClickHouse client to create a ClickHouse user.

Step 4 Create the **test001** database.

create database test001 on cluster default_cluster;

Step 5 Create the replication table **test010**.

CREATE TABLE *test001.test010* on cluster default_cluster (

`EventDate` DateTime, `CounterID` UInt32, `UserID` UInt32, `ver` UInt16) ENGINE = ReplicatedReplacingMergeTree('/clickhouse/tables/{shard}/test3', '{replica}', ver) ORDER BY (CounterID, EventDate, intHash32(UserID));

- **Step 6** Create a replication table **test010** and insert data into the table. insert into *test001.test010* values('2020-01-29',111,111,111);
- Step 7 View data in the test010 table.
 select * from test001.test010;

Figure 4-5 Viewing data in the test010 table.

EventDate	-Counter ID	User ID	ver
2020-01-29 00:00:00	111	111	111
1 rows in set. Elapsed:	0.003 sec.		

- Step 8 Create a distributed table test001.test010_dir and insert data into the table. create table test001.test010_dir ON CLUSTER default_cluster as test001.test010 ENGINE = Distributed(default_cluster, test001, test010, rand()); insert into test001.test010_dir values('2020-01-29',111,111,111);
- **Step 9** Query the distributed table.

select * from test001.test010_dir;

Figure 4-6 Viewing data in the test010_dir table

EventDate	-Counter ID	UserID	ver
2020-01-29 00:00:00	111	111	
l rows in set. Elapsed:	0.003 sec.		

Step 10 Delete the created database table. drop database *test001* on cluster default_cluster no delay;

----End

Follow-up Operations: Releasing Resources

To avoid additional expenditures, release resources promptly if you no longer need them. For details, see **Deleting an MRS Cluster**.

Related Information

For details about how to manage ClickHouse permissions, import data from RDS for MySQL, OBS, HDFS, and GaussDB(DWS) to ClickHouse tables, manage multiple ClickHouse tenants, and access ClickHouse through ELB, see Using ClickHouse.

5 Creating and Using an MRS Cluster Requiring Security Authentication

Scenario

This topic helps you create a Hadoop analysis cluster that requires Kerberos authentication and submit a wordcount job through the cluster client. A wordcount job is a classic Hadoop job that counts words in massive amounts of text.

The Hadoop cluster uses many open-source Hadoop ecosystem components, including YARN for cluster resource management and Hive and Spark for offline large-scale distributed data storage and compute to provide massive data analysis and query capabilities.

Procedure

Before you start, complete operations described in **Preparations**. Then, follow these steps:

- 1. **Creating an MRS Cluster**: Create a Hadoop analysis cluster of MRS 3.2.0-LTS.1 that requires Kerberos authentication.
- 2. **Creating a Cluster User**: Create a role that has the permission to submit the wordcount job and bind the role to a user on FusionInsight Manager.
- 3. Installing the Cluster Client: Download and install the MRS cluster client.
- 4. **Preparing Applications and Data**: Prepare the data files required for running the wordcount sample program on the MRS cluster client.
- 5. **Submitting a Job and Viewing the Result**: Submit a wordcount data analysis job on the cluster client and view the execution result.
- 6. **Releasing resources**: To avoid additional expenditures, release resources promptly if you no longer need them.

Preparations

• You have prepared an IAM user who has the permission to create MRS clusters. For details, see **Creating an MRS User**.

Step 1: Creating an MRS Cluster

- **Step 1** Go to the **Buy Cluster** page.
- **Step 2** Search for MapReduce Service in the service list and enter the MRS console.
- **Step 3** Click Buy Cluster. The **Quick Config** tab is displayed.
- **Step 4** Configure the cluster as you need. In this example, a pay-per-use MRS 3.2.0-LTS.1 cluster will be created. For more details about how to configure the parameters, see **Quickly Creating a Cluster**.

 Table 5-1 MRS cluster parameters

Parameter	Example Value	Description
Billing Mode	Pay-per-use	Billing mode of the cluster you want to create. MRS provides two billing modes: yearly/monthly and pay-per-use.
		Pay-per-use is a postpaid billing mode. You pay as you go and pay for what you use. The cluster usage is calculated by the second but billed every hour.
Region	CN-Hong Kong	Region where the MRS resources to be requested belong.
		MRS clusters in different regions cannot communicate with each other over an intranet. For lower network latency and quick resource access, select the nearest region.
Cluster Name	mrs_demo	Name of the MRS cluster you want to create.
Cluster Type	Custom	A range of clusters that accommodate diverse big data demands. You can select a Custom cluster to run a wide range of analytics components supported by MRS.
Version Type	LTS	Service type of the MRS
Cluster Version	MRS 3.2.0- LTS.1	Version of the MRS cluster. Supported open-source components and their functions vary depending on the cluster version. You are advised to select the latest version.
Component	Hadoop Analysis Cluster	Cluster templates containing preset opensource components you will need for your business.
AZ	AZ 1	Available AZ associated with the cluster region.
VPC	vpc-default	VPC where you want to create the cluster. You can click View VPC to view the name and ID. If no VPC is available, create one.

Parameter	Example Value	Description
Subnet	subnet- default	Subnet where your cluster belongs. You can access the VPC management console to view the names and IDs of existing subnets in the VPC. If no subnet is created under the VPC, click Create Subnet to create one.
Cluster Node	Default value	Cluster node details.
Kerberos Authenticat ion	Enabled	Whether Kerberos authentication is enabled.
Username	admin/root	Username for logging in to the cluster management page and the ECS node.
Password	-	User password for logging in to the cluster management page and the ECS node.
Confirm Password	-	Enter the user password again.
Enterprise Project	default	Enterprise project to which the cluster belongs.
Secure Communic ations	Selected	Select the check box to agree to use the access control rules.

Figure 5-1 Buying a Hadoop analysis cluster

Cluster Configuration			
Cluster Name			
ms			
The name must be unique. It can contain only 1 to 64 characters. Only letters, digits	; hyphens (-), and underscores (_) are allowed.		
Cluster Type			
Custom The periods a such range of customcable component controls. Composed management toget dia demands help you bey clusters fastler.	Analysis cluster • Guidate for analysing and processing massive amounts of offline data to drain need to an usually require large scale compute and intronge macroscillates usually require large scale compute and cluster analysis components, such as Haddog, Spark, Hillace, Hine, Flink, Outrie and Tot are required.	Streaming cluster I lead for quick analysis of read-time data sources. Setteming data sources. Setteming data sources usually has high CPU and memory requirements. Components such as Kalta and Flume are recommended.	Hybrid cluster • Solialise for both offline data analysis and shream processing • You can select waturbaries components such as Holicone, Space, Holizo, and Hive, and shream processing components such as Katha and Plume.
Vention Type (*) Normal Long Tem Boycon (137): employs MRTS: own components to provide highly reliab Cong Venter Vention MR3.32.8.413.1 v Component	is clusters with alrang DR capabitities, making long-lerm support and evolution possible		
Real-time Analysis Cluster Massive data celection, real-time data analysis and query Haddop 3.11 Filek 1.15.0 Katha 2.112.4.0 ZooKkeeper 3.6.3 Ranger 2.00 ClickHouse 22.3.2	ClickHouse Cluster A Colum Database Management System (DBMS) for On-Line An ZooKeeper 1.6.1 ClubHouse 22.1.2.2	Hadoop Analysis Cluster Analysis and query of veral amounts of data Hadoop 3.11 Heve 3.10 Spart/24.31.1 Pink 115.0 Zookkeeper 3.6.3 Ranger 2.0.0 Riz 0.9.2	HBase Query Cluster Massive data storage and millioscond-level data quaries Haddop 3.3.1 HBase 22.3 ZooKeeper 3.6.3 Ranger 2.0.0

- **Step 5** Click Buy Now. A page is displayed showing that the task has been submitted.
- **Step 6** Click **Back to Cluster List**. You can view the status of the newly created cluster on the **Active Clusters** page.

Wait for the cluster creation to complete. The initial status of the cluster is **Starting**. After the cluster is created, the cluster status becomes **Running**.

----End

Step 2: Creating a Cluster User

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

- **Step 1** Click the MRS cluster name in the cluster list to go to the dashboard page.
- **Step 2** Click **Access Manager** next to **MRS Manager**. In the displayed dialog box, select **EIP** and configure the EIP information.

For the first access, click **Manage EIPs** to purchase an EIP on the EIP console. Go back to the **Access MRS Manager** dialog box, refresh the EIP list, and select the EIP.

Step 3 Select the confirmation check box and click **OK** to log in to the FusionInsight Manager of the cluster.

The username for logging in to FusionInsight Manager is **admin**, and the password is the one configured during cluster purchase.

- **Step 4** Click **System** in the navigation pane on the top, and click **Permission** > **Role**.
- **Step 5** Click **Create Role** and set the following parameters. For details, see **Creating a Role**.
 - Enter a role name, for example, **mrrole**.
 - For Configure Resource Permission, select the cluster to be operated, choose Yarn > Scheduler Queue > root, and select Submit and Admin in the Permission column. Click the name of the target cluster in the path and then configure other permissions.

Figure 5-2 Configuring resource permissions for YARN

Configure Resource Permission:	n. All resources • mrs Yam • Scheduler Queue • root				
	D	D	Permi	ission	
	Resource Name	Resource Type	Submit	Admin	
	launcher-job	Leaf Queue	☑ ③	I (1)	
	default	Leaf Queue	☑ ①	•	

 Choose HDFS > File System > hdfs://hacluster/. Locate the row that contains user, select Read, Write, and Execute in the Permission column, and click OK.

Figure 5-3 Configuring resource permissions for HDFS

Configure Resource Permission:	All resources HDFS > File S	ystem						
	Resource Name	Resource Type	Permission					
	Resource Name	Resource Type	Read	Vite Vite	Execute			
	hdfs://hacluster/	Folder	I	I	③			
	viewfs://ClusterX/	Folder						

- **Step 6** Click **User** in the navigation pane on the left, and then click **Create** on the displayed page. Set the following parameters. For details, see **Creating a User**.
 - Enter a username, for example, **test**.
 - Set User Type to Human-Machine.
 - Enter the password in **Password** and enter it again in **Confirm Password**.
 - Bind Manager_viewer to the mrrole role created in Step 5 to grant permissions.

Figure 5-4	Creating a user	
* Username:	test ×	
★ User Type:	Human-Machine Machine-Machine	
* Password Policy:	default Select	
* Password:		
* Confirm Password:		
User Group:	Add Clear All Create User Group	
Primary Group:		
Role:	Add Clear All Create Role	
	mrrole × Manager_viewer ×	

Step 7 Click OK.

----End

Step 3: Installing the Cluster Client

You need to install a cluster client to connect to component services in the cluster and submit jobs.

You can install the clients on a node in or outside the cluster. This topic installs the client on the **Master1** node as an example.

- **Step 1** Click the MRS cluster name in the cluster list to go to the dashboard page.
- **Step 2** Click **Access Manager** next to **MRS Manager**. In the displayed dialog box, select **EIP** and configure the EIP information.

For the first access, click **Manage EIPs** to purchase an EIP on the EIP console. Go back to the **Access MRS Manager** dialog box, refresh the EIP list, and select the EIP.

Step 3 Select the confirmation check box and click **OK** to log in to the FusionInsight Manager of the cluster.

The username for logging in to FusionInsight Manager is **admin**, and the password is the one configured during cluster purchase.

Step 4 On the displayed **Homepage** page, click •••• next to the cluster name and click **Download Client** to download the cluster client.

Figure 5-5 Downloading the client

Cluster

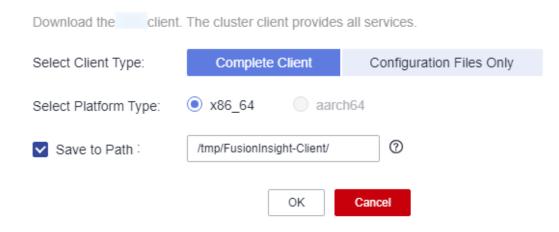
Test MRS	≈		
≡ 3 Hosts	Start		
	Stop		
🔍 🗟 DBServio	Restart		
• H HBase	Rolling-restart Service		
• • • • • • • • • • • • • • • • • • •	Synchronize		
🖲 🦪 HDFS	Configurations		
	Restart Configuration- Expired Instances		
🔍 😺 Hive	Health Check		
🔍 🏕 KrbServe	Download Client		
• 🌒 LdapSer	Export Installation Template		

In the **Download Cluster Client** dialog box, set the following parameters:

- Set Select Client Type to Complete Client.
- Retain the default value for **Platform Type**, for example, **x86_64**.
- Retain the default path for **Save to Path**. The generated file will be saved in the **/tmp/FusionInsight-Client** directory on the active OMS node of the cluster.

Figure 5-6 Downloading the cluster client

Download Cluster Client



Click **OK** and wait until the client software is generated.

Step 5 Go back to the MRS console and click the cluster name in the cluster list. Go to the Nodes tab, click the name of the node that contains master1. In the upper right corner of the ECS details page, click Remote Login to log in to the Master1 node.

Service ZooKeeper / Instances			
Service Status Instances Service	Configuration		
Add Instance More ∨)		
Role	Host Name \ominus	OM IP Address	Business IP Address
quorumpeer	node-master2ggzl.mrs-82mt.com	192.1	192. ⁻
quorumpeer	node-master3PywB.mrs-82mt.com	192.1	192
quorumpeer	node-master1cwUo.mrs-82mt.com	192.1	192.1

Figure 5-7 Checking the Master1 node

- **Step 6** Log in to the **Master1** node as user **root**. The password is the one you set for the **root** user during cluster purchase.
- **Step 7** Switch to the directory where the client software package is stored and decompress the package.

cd /tmp/FusionInsight-Client/

tar -xvf FusionInsight_Cluster_1_Services_Client.tar

tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar

Step 8 Go to the directory where the installation package is stored and install the client.

cd FusionInsight_Cluster_1_Services_ClientConfig

Install the client to a specified directory (an absolute path), for example, **/opt/ client**.

./install.sh /opt/client

... component client is installed successfully

NOTE

A client installation directory will be automatically created if it does not exist. If there is such directory, it must be empty. The directory name cannot contain spaces. The client installation directory can contain only uppercase letters, lowercase letters, digits, and underscores (_).

----End

Step 4: Preparing Applications and Data

You can run the wordcount sample program preset in the cluster client on the created cluster, or develop a big data application and upload it to the cluster. This topic uses the wordcount sample program preset in the cluster client as an example. You need to prepare the data files required for running the wordcount sample program.

- **Step 1** Log in to the **Master1** node as user **root**.
- **Step 2** Prepare data files.

There is no format requirement. For example, the file names are **wordcount1.txt** and **wordcount2.txt**, and the content is as follows:

vi /opt/wordcount1.txt

hello word hello wordcount

vi /opt/wordcount2.txt

hello mapreduce hello hadoop

Step 3 Switch to the client installation directory, configure environment variables, and create an HDFS directory for storing sample data, for example, /user/example/ input.

cd /opt/client

source bigdata_env

kinit test (**test** is the username created in **Step 6**. Change the password upon first login.)

hdfs dfs -mkdir /user/example

hdfs dfs -mkdir /user/example/input

NOTE

The **test** user created in **Step 6** has only the read, write, and execute permissions on the / **user** directory. If the **input** directory is created in a directory other than **/user**, an error message is displayed, indicating that the permission is required. The following is an example:

hdfs dfs -mkdir /hbase/input

The following error message is displayed:

mkdir: Permission denied: user=test, access=EXECUTE, inode="/hbase":hbase:hadoop:drwxrwx--T

Step 4 Upload the sample data to HDFS.

hdfs dfs -put /opt/wordcount1.txt /user/example/input

hdfs dfs -put /opt/wordcount2.txt /user/example/input

----End

Step 5: Submitting a Job and Viewing the Result

- **Step 1** Log in to the client node (Master1) as user root.
- **Step 2** Submit the wordcount job, read source data for analysis, and output the execution result to the HDFS.

cd /opt/client

source bigdata_env

kinit test

hadoop jar HDFS/hadoop/share/hadoop/mapreduce/hadoop-mapreduceexamples-3.3.1-*.jar wordcount "/user/example/input/*" "/user/example/ output/"

File Input Format Counters Bytes Read=56 File Output Format Counters Bytes Written=48

D NOTE

- /user/example/output/ indicates the address for storing job output files on the HDFS. Set it to a directory that does not exist.
- The name of the **hadoop-mapreduce-examples-3.3.1-*.jar** file varies depending on the cluster client version. Use the actual name.

Step 3 Query job execution results.

1. View the job output file.

hdfs dfs -ls /user/example/output/

... ... /user/example/output/_SUCCESS

- ... /user/example/output/part-r-0000
- 2. Save the output in the HDFS file system. You can run a command to download the output to the local PC and view it.

The following is an example:

hdfs dfs -get /user/example/output/part-r-00000 /opt

cat /opt/part-r-00000

The content of the part-r-00000 file is as follows:

```
hadoop 1
hello 4
mapreduce 1
word 1
wordcount 1
```

Step 4 View job run logs.

- Log in to FusionInsight Manager of the cluster as user test created in Step 6 and choose Cluster > Services > Yarn.
- 2. Click the **ResourceManager(xxx,Active)** link in the row where **ResourceManager Web UI** is.
- 3. On the **All Applications** page, click the ID of the target job to view the job details.

NOTE

On the **All Applications** page, you can confirm a task based on the task submission time and the user name that submits the task.

												All Ap	prica	ciona							
uster	Cluster Met																				
Labels	Apps	Submitted	Apps	Pending	Apps P	unning	Apps Comp	leted	Contai	ners Running		Used Resour	ces		Total Ro				ved Resources		Phy
- halo	2		0		0	2			0		<memory.< td=""><td>0.8, vCores:0></td><td></td><td><memory< td=""><td>c48 GB, vCores:2</td><td>4></td><td>10</td><td>remory:0 B, vCo</td><td>restor</td><td></td><td>87</td></memory<></td></memory.<>	0.8, vCores:0>		<memory< td=""><td>c48 GB, vCores:2</td><td>4></td><td>10</td><td>remory:0 B, vCo</td><td>restor</td><td></td><td>87</td></memory<>	c48 GB, vCores:2	4>	10	remory:0 B, vCo	restor		87
	Cluster Nod																				
		Active Nodes			Decc	mmissioning Noo	les			Decomm	issioned Nodes			Lost Node			Unhealthy No	des		Reb	ooted Noder
MITTED	2			2					Q				2		\$				2		
A SAVING MITTED	User Metric	s for test																			
15HED	Apps 5	ubmitted	Apps Pe	nding	Apps Runnis	a As	ops Completed		Containers Runni	ng	Containers Pend	áng	Containe	n Reserved	Memor		Memory Pe			Reserved	VCo
	1		0	0		1		0		0		0			0.8	0	8		0.8		0
isc	Scheduler N	detrics																			
80	Sche	duler Type		Scheduling	Resource Typ	10	Minimu	m Allocation	1		Maximu	m Allocation			Maximu	m Cluster Appli	cation Priority	Sche	daler Busy %	RM D	ispatcher Eve
	SuperiorYarr	Scheduler	(yamio/gp	u, memory-mb ().	unit+Mi), vco	es)	<memory 512,<="" td=""><td>vCores:1></td><td><memo< td=""><td>ry/65536, vCore</td><td>s 32, yam io/gpu: 5</td><td>2233720368547</td><td>75807></td><td></td><td>5</td><td></td><td></td><td>0</td><td></td><td>0</td><td></td></memo<></td></memory>	vCores:1>	<memo< td=""><td>ry/65536, vCore</td><td>s 32, yam io/gpu: 5</td><td>2233720368547</td><td>75807></td><td></td><td>5</td><td></td><td></td><td>0</td><td></td><td>0</td><td></td></memo<>	ry/65536, vCore	s 32, yam io/gpu: 5	2233720368547	75807>		5			0		0	
	Show 20 ¥	entries																			
		ID	• User	QueueUser	Name	Application Type	Application Tags	Queue	Application Priority	StortTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Allocated GPUs	Reserved CPU VCores	Reserved Memory MB	Reserve
				test		MAPREDUCE		default		Thu Jun 27	Thu Jun 27	Thu Jun 27	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A	N/A	N/A

----End

Follow-up Operations: Releasing Resources

To avoid additional expenditures, release resources promptly if you no longer need them. For details, see **Deleting an MRS Cluster**.

Related Information

Hadoop components include HDFS, YARN, and MapReduce. You can run jobs to analyze or view offline data. For details, see **Using HDFS**, **Using MapReduce**, or **Using YARN**.

6 Best Practices for Beginners

After an MRS cluster is deployed, you can try some practices provided by MRS to meet your service requirements.

Practice		Description
Data analytics	Using Spark2x to Analyze IoV Drivers' Driving Behavior	This practice describes how to use Spark to analyze driving behavior. You can get familiar with basic functions of MRS by using the Spark2x component to analyze and collect statistics on driving behavior, obtain the analysis result, and collect statistics on the number of violations such as sudden acceleration and deceleration, coasting, speeding, and fatigue driving in a specified period.
	Using Hive to Load HDFS Data and Analyze Book Scores	This practice describes how to use Hive to import and analyze raw data and how to build elastic and affordable offline big data analytics. In this practice, reading comments from the background of a book website are used as the raw data. After the data is imported to a Hive table, you can run SQL commands to query the most popular best- selling books.
	Using Hive to Load OBS Data and Analyze Enterprise Employee Information	This practice describes how to use Hive to import and analyze raw data from OBS and how to build elastic and affordable big data analytics based on decoupled storage and compute resources. This practice describes how to develop a Hive data analysis application and how to run HQL statements to access Hive data stored in OBS after you connect to Hive through the client. For example, manage and query enterprise employee information.

Table 6-1 Best practices

Practice		Description
	Using Flink Jobs to Process OBS Data	This practice describes how to use the built-in Flink WordCount program of an MRS cluster to analyze the source data stored in the OBS file system and calculate the number of occurrences of specified words in the data source. MRS supports decoupled storage and compute in scenarios where a large storage capacity is required and compute resources need to be scaled on demand. This allows you to store your data in OBS and use an MRS cluster only for data computing.
Data migratio n	Data Migration Solution	This practice describes how to migrate HDFS, HBase, and Hive data to an MRS cluster in different scenarios. You will try to prepare for data migration, export metadata, copy data, and restore data.
	Migrating Data from Hadoop to MRS	In this practice, CDM is used to migrate data (dozens of terabytes or less) from Hadoop clusters to MRS.
	Migrating Data from HBase to MRS	In this practice, CDM is used to migrate data (dozens of terabytes or less) from HBase clusters to MRS. HBase stores data in HDFS, including HFile and WAL files. The hbase.rootdir configuration item specifies the HDFS path. By default, data is stored in the /hbase folder on MRS.
		Some mechanisms and tool commands of HBase can also be used to migrate data. For example, you can migrate data by exporting snapshots, exporting/importing data, and CopyTable.
	Migrating Data from Hive to MRS	In this practice, CDM is used to migrate data (dozens of terabytes or less) from Hive clusters to MRS.
		 Hive data migration consists of two parts: Hive metadata, which is stored in the databases such as MySQL. By default, the metadata of the MRS Hive cluster is stored in MRS DBService (Huawei GaussDB database). You can also use RDS for MySQL as the external metadata database.
		 Hive service data, which is stored in HDFS or OBS

Practice		Description			
	Migrating Data from MySQL to an MRS Hive	This practice demonstrates how to use CDM to import MySQL data to the Hive partition table in an MRS cluster.			
	Partitioned Table	Hive supports SQL to help you perform extraction, transformation, and loading (ETL) operations on large-scale data sets. Queries on large-scale data sets take a long time. In many scenarios, you can create Hive partitions to reduce the total amount of data to be scanned each time. This significantly improves query performance.			
	Migrating Data from MRS HDFS to OBS	This practice demonstrates how to migrate file data from MRS HDFS to OBS using CDM.			
System Intercon nection	Using DBeaver to Access Phoenix	This practice describes how to use DBeaver to access Phoenix. The local DBeaver can connect to the HBase component in the MRS cluster through the Phoenix Jar package. After they are connected, you can create an HBase table and insert data into the table using DBeaver.			
	Using DBeaver to Access HetuEngine	This practice describes how to use DBeaver to access HetuEngine. The local DBeaver can connect to the HetuEngine component in the MRS cluster through the JDBC Jar package. After they are connected, you can view information about the data sources connected to HetuEngine with DBeaver.			
	Interconnecting Hive with External Self- Built Relational Databases	This practice describes how to use Hive to connect to open-source MySQL and Postgres databases. After an external metadata database is deployed in a cluster that has Hive data, the original metadata tables will not be automatically synchronized. Before installing Hive, determine whether to store metadata in an external database or DBService. For the former, deploy an external database when installing Hive or when there is no Hive data. After Hive installation, the metadata storage location cannot be changed. Otherwise, the original metadata will be lost.			

Practice		Description
	Interconnecting Hive with CSS	This practice describes how to use Hive to interconnect with CSS Elasticsearch.
		In this practice, you will use the Elasticsearch- Hadoop plug-in to exchange data between Hive and Elasticsearch of Cloud Search Service (CSS) so that Elasticsearch index data can be mapped to Hive tables.