

# MapReduce 服务

## 快速入门

文档版本 01  
发布日期 2024-08-03



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

## 商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

## 注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

# 华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

---

## 目录

---

1 快速创建和使用 Hadoop 离线数据分析集群.....	1
2 快速创建和使用 Kafka 流式数据处理集群.....	8
3 快速创建和使用 HBase 离线数据查询集群.....	16
4 快速创建和使用 ClickHouse 列式数据库集群.....	23
5 快速创建和使用启用安全认证的 MRS 集群.....	30
6 入门实践.....	40

# 1 快速创建和使用 Hadoop 离线数据分析集群

## 操作场景

本入门提供从零开始创建Hadoop离线数据分析集群并通过集群客户端提交一个wordcount作业的操作指导。wordcount是最经典的Hadoop作业，用于统计海量文本的单词数量。

Hadoop集群完全使用开源Hadoop生态，采用Yarn管理集群资源，提供Hive、Spark离线大规模分布式数据存储和计算及进行海量数据分析与查询的能力。

## 操作流程

开始使用如下样例前，请务必按[准备工作](#)指导完成必要操作。

1. **步骤一：创建MRS集群**：创建一个MRS 3.1.5版本的“Hadoop分析集群”。
2. **步骤二：安装集群客户端**：下载并安装MRS集群客户端。
3. **步骤三：准备应用程序及数据**：准备MRS集群客户端内wordcount样例程序运行所需的数据文件。
4. **步骤四：提交作业并查看结果**：在集群客户端提交wordcount数据分析作业并查看执行结果。

## 准备工作

- 注册账号并实名认证。  
在创建MRS集群之前，请先[注册华为账号并开通华为云](#)，进行[实名认证](#)。  
如果您已开通华为云并进行实名认证，请忽略此步骤。
- 已准备具有创建MRS集群的IAM用户，详细操作请参见[创建IAM用户并授权使用MRS](#)。

## 步骤一：创建 MRS 集群

**步骤1** 进入[购买MRS集群页面](#)。

**步骤2** 在服务列表中搜索“MapReduce服务 MRS”，进入MRS服务管理控制台。

**步骤3** 单击“购买集群”，进入“购买集群”页面，选择“快速购买”页签。

**步骤4** 根据实际业务规划情况填写集群配置信息（本示例为快速创建按需计费的MRS 3.1.5版本集群，如需了解更多参数配置请参考[快速创建MRS集群](#)）。

表 1-1 MRS 集群配置参数

参数名称	参数说明	取值样例
计费模式	选择待创建的集群的计费模式，MRS提供“包年/包月”与“按需计费”两种计费模式。 按需计费是一种后付费模式，即先使用再付费，按照MRS集群实际使用时长计费。	按需计费
区域	选择区域。 不同区域的云服务产品之间内网互不相通。请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。	中国-香港
集群名称	待创建的MRS集群名称。	mrs_demo
集群类型	根据实际业务需要选择待创建的MRS集群类型。 “自定义”类型集群提供丰富的组件搭配，可自行选择对应版本MRS集群所支持的所有组件。	自定义
版本类型	待创建的MRS集群版本类型。	普通版
集群版本	待创建的MRS集群版本，不同版本所包含的开源组件版本及功能特性可能不同，推荐选择最新版本。	MRS 3.1.5
组件选择	基于系统预置的集群模板选择要购买的集群组件。	Hadoop分析集群
可用区	选择集群工作区域下关联的可用区。	可用区1
虚拟私有云	选择需要创建集群的VPC，单击“查看虚拟私有云”进入VPC服务查看已创建的VPC名称和ID。如果没有VPC，需要创建一个新的VPC。	vpc-default
子网	选择需要创建集群的子网，可进入VPC服务查看VPC下已创建的子网名称和ID。如果VPC下未创建子网，请单击“创建子网”进行创建。	subnet-default
集群节点	配置集群节点信息。	保持默认
Kerberos 认证	是否启用Kerberos认证。	不开启
用户名	登录集群管理页面及ECS节点的用户名。	admin/root
密码	设置登录集群管理页面及ECS节点用户的密码。	-
确认密码	再次输入设置用户密码。	-
企业项目	选择集群所属的企业项目。	default
通信安全授权	勾选确认授权。	勾选

图 1-1 购买 Hadoop 分析集群

计费模式 ? 包年/包月 按需计费

集群名称

集群类型 自定义 更多类型

自定义集群

- 提供丰富的组件搭配，可自行选择对应版本MRS集群所支持的所有组件。
- 可按业务需求自定义调整集群服务的部署方式，实现管理控制角色分开部署、管理控制角色合并部署或数据角色合并部署。
- 建议不要将多个数据存储类服务部署在一个节点组中，避免资源争抢。

版本类型 ? LTS版 普通版

集群版本

组件选择

实时分析集群	ClickHouse集群	Hadoop分析集群	HBase查询集群
Hadoop 3.1.1, Flink 1.12.2, Kafka 2.11-2.4.0, ZooKeeper 3.6.3, Ranger 2.0.0 and ClickHouse 21.3.4.25	ZooKeeper 3.6.3 and ClickHouse 21.3.4.25	Hadoop 3.1.1, Hive 3.1.0, Spark2x 3.1.1, Tez 0.9.2, Flink 1.12.2, ZooKeeper 3.6.3, Ranger 2.0.0 and Presto 333	Hadoop 3.1.1, HBase 2.2.3, ZooKeeper 3.6.3 and Ranger 2.0.0
海量数据采集，数据实时分析与查询	用于联机分析(OLAP)的列式数据库管理系统(DBMS)	海量数据分析与查询	海量数据存储，实现毫秒级数据查询

**步骤5** 单击“立即购买”，进入任务提交成功页面。

**步骤6** 单击“返回集群列表”，在“现有集群”列表中可以查看到集群创建的状态。

集群创建需要时间，所创集群的初始状态为“启动中”，创建成功后状态更新为“运行中”，请您耐心等待。

---结束

## 步骤二：安装集群客户端

MRS集群创建成功后，用户需手动安装集群客户端用于连接集群内各组件服务，进行客户端Shell连接、作业任务提交等操作。

客户端可以安装在集群内的节点上，也可以安装在集群外的节点上，本指南以在集群内的Master1节点上安装客户端为例进行介绍。

**步骤1** MRS集群创建成功后，在集群列表中单击MRS集群名称进入集群概览页面。

**步骤2** 单击“集群管理页面”后的“前往 Manager”，在弹出的窗口中选择“EIP访问”并配置弹性IP信息。

首次访问时，可点击“管理弹性公网IP”，在弹性公网IP控制台购买一个弹性公网IP，购买成功后刷新弹性公网IP列表并选择。

**步骤3** 勾选确认信息后，单击“确定”，登录集群的FusionInsight Manager管理界面。

Manager登录用户名为**admin**，密码为购买集群时配置的用于登录Manager管理界面的“admin”用户的自定义密码。

**步骤4** 在“主页”页签的集群名称后单击 ⋮，单击“下载客户端”下载集群客户端。

图 1-2 下载客户端  
集群



在“下载集群客户端”弹窗中，配置以下参数：

- 选择客户端类型：选择“完整客户端”。
- 选择平台类型：必须与待安装客户端的节点的架构匹配，例如“x86\_64”。对于集群内的节点，节点架构类型可在Manager页面选择“主机”，在待安装客户端节点所在行的“平台类型”列查看。
- 勾选“仅保存到如下路径”，使用默认保存路径，文件生成后将保存在集群主 OMS 节点（通常为 Master1 节点）的“/tmp/FusionInsight-Client”目录下。

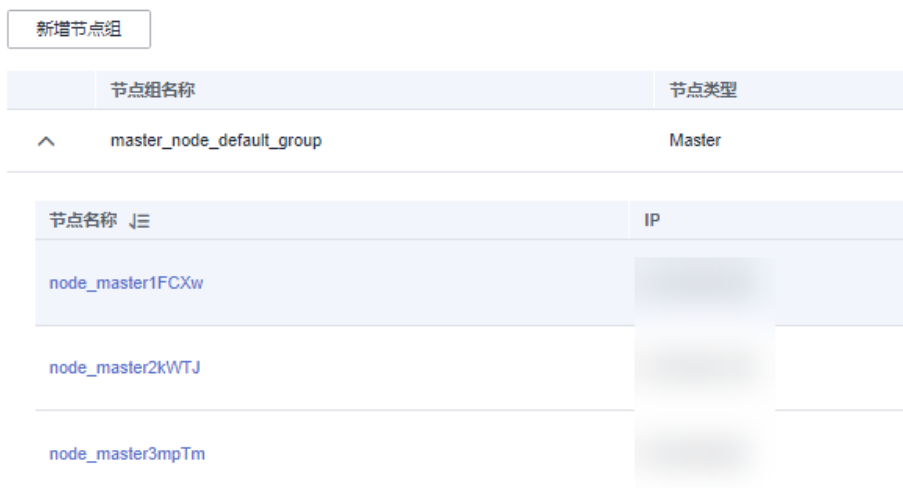
图 1-3 下载集群客户端提示框



单击“确定”后，等待客户端软件生成成功。

**步骤5** 在MRS服务管理控制台的集群列表中，单击集群名称，在集群的“节点管理”页签，单击包含“master1”的节点名称，在ECS详情页面右上角单击“远程登录”登录Master1节点。

图 1-4 查看 Master1 节点



**步骤6** 使用root用户登录Master1节点，密码为购买集群时配置用于登录集群节点的“root”用户的自定义密码。

**步骤7** 执行以下命令切换到客户端软件包所在目录，并解压客户端软件包。

```
cd /tmp/FusionInsight-Client/  
tar -xvf FusionInsight_Cluster_1_Services_Client.tar  
tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar
```

**步骤8** 执行以下命令进入安装包所在目录，安装客户端。

```
cd FusionInsight_Cluster_1_Services_ClientConfig
```

执行如下命令安装客户端到指定目录（目录可以不存在，如果存在则必须为空）。

例如安装到“/opt/client”目录，执行以下命令：

```
./install.sh /opt/client
```

等待客户端安装完成。

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

----结束

### 步骤三：准备应用程序及数据

MRS集群创建成功后，可获取集群客户端内预置的wordcount样例程序进行运行，也可准备自行开发的大数据应用程序并上传至集群。

本指南以运行MRS集群客户端内的wordcount样例程序为例进行介绍，需先准备wordcount样例程序运行所需的数据文件。



**步骤1** 以root用户登录Master1节点。

**步骤2** 准备数据文件。

例如文件名为“wordcount1.txt”和“wordcount2.txt”，内容如下所示：

**vi /opt/wordcount1.txt**

```
hello word  
hello wordcount
```

**vi /opt/wordcount2.txt**

```
hello mapreduce  
hello hadoop
```

**步骤3** 执行以下命令切换到客户端安装目录，配置环境变量并创建用于存放样例数据的HDFS目录，例如“/user/example/input”。

**cd /opt/client**

**source bigdata\_env**

**hdfs dfs -mkdir /user/example/input**

**步骤4** 执行以下命令，将样例数据上传至HDFS。

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

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

----结束

## 步骤四：提交作业并查看结果

**步骤1** 以root用户登录客户端节点（Master1节点）。

**步骤2** 执行如下命令提交wordcount作业，读取源数据进行分析并将执行结果输出至HDFS。

**cd /opt/client**

**source bigdata\_env**

**hadoop jar HDFS/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-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
```

### 📖 说明

- “/user/example/output/”为HDFS上存放作业输出文件地址，请设置为一个不存在的目录。
- “hadoop-mapreduce-examples-3.3.1-\*.jar”文件在不同版本的集群客户端目录下文件名有差异，以实际名称为准。

**步骤3** 查看作业执行结果。

- 作业执行完成后，可执行以下命令查看作业输出文件。

**hdfs dfs -ls /user/example/output/**

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

2. 输出结果保存在HDFS文件系统中，可以通过命令将其下载到本地进行查看。  
例如执行以下命令：

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

“part-r-00000” 文件内容如下：

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

#### 步骤4 查看作业运行记录。

1. 使用admin用户登录运行作业的集群的FusionInsight Manager页面，选择“集群 > 服务 > Yarn”。
2. 单击ResourceManager Web UI所在行的“ResourceManager(xxx, 主)”链接。
3. 在“All Applications”页面，单击目标作业ID，可以查看作业运行的详情。

#### 📖 说明

在“All Applications”页面可根据任务提交时间和提交任务的用户名确认具体任务。

图 1-5 查看作业详情

ID	User	Queue	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU V-Cores	Allocated Memory MB	Allocated GPUs	Reserved CPU V-Cores	Reserved Memory MB	Reserved GPUs	% Quota	
application_171627489965_0008	root	root	wordcount	MAPREDUCE	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	N/A	0.0

----结束

## 相关信息

Hadoop组件包含了HDFS、Yarn、Mapreduce服务，可根据业务实际需求运行相关作业进行离线数据分析或查看，详细操作请参见[使用HDFS](#)、[使用MapReduce](#)或[使用Yarn](#)。

# 2 快速创建和使用 Kafka 流式数据处理集群

## 操作场景

本入门提供从零开始创建流式分析集群并在Kafka主题中产生和消费消息的操作指导。

Kafka集群提供一个高吞吐量、可扩展性的消息系统，广泛用于日志收集、监控数据聚合等场景。Kafka可实现高效的流式数据采集、实时数据处理存储等。

## 操作流程

开始使用如下样例前，请务必按[准备工作](#)指导完成必要操作。

1. **步骤一：创建MRS集群**：创建一个MRS 3.2.0-LTS.1版本的“实时分析集群”。
2. **步骤二：安装集群客户端**：下载并安装MRS集群客户端。
3. **步骤三：使用Kafka客户端创建Topic**：在Kafka客户端创建Topic。
4. **步骤四：管理Kafka主题中的消息**：在Kafka客户端消费已创建的Topic中的消息。

## 准备工作

- 注册账号并实名认证。  
在创建MRS集群之前，请先[注册华为账号并开通华为云](#)，进行[实名认证](#)。  
如果您已开通华为云并进行实名认证，请忽略此步骤。
- 已准备具有创建MRS集群的IAM用户，详细操作请参见[创建IAM用户并授权使用MRS](#)。

## 步骤一：创建 MRS 集群

- 步骤1** 进入[购买MRS集群页面](#)。
- 步骤2** 在服务列表中搜索“MapReduce服务 MRS”，进入MRS服务管理控制台。
- 步骤3** 单击“购买集群”，进入“购买集群”页面，选择“快速购买”页签。
- 步骤4** 根据实际业务规划情况填写集群配置信息（本示例为快速创建按需计费的MRS 3.2.0-LTS.1版本集群，如需了解更多参数配置请参考[快速创建MRS集群](#)）。

表 2-1 MRS 集群配置参数

参数名称	参数说明	取值样例
计费模式	选择待创建的集群的计费模式，MRS提供“包年/包月”与“按需计费”两种计费模式。 按需计费是一种后付费模式，即先使用再付费，按照MRS集群实际使用时长计费。	按需计费
区域	选择区域。 不同区域的云服务产品之间内网互不相通。请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。	中国-香港
集群名称	待创建的MRS集群名称。	mrs_demo
集群类型	根据业务实际需要选择待创建的MRS集群类型。“自定义”类型集群提供丰富的组件搭配，可自行选择对应版本MRS集群所支持的所有组件。	自定义
版本类型	待创建的MRS集群版本类型，不同版本所包含的开源组件版本及功能特性可能不同，推荐选择最新版本。	LTS版
集群版本	待创建的MRS集群版本。	MRS 3.2.0-LTS.1
组件选择	基于系统预置的集群模板选择要购买的集群组件。	实时分析集群
可用区	选择集群工作区域下关联的可用区。	可用区1
虚拟私有云	选择需要创建集群的VPC，单击“查看虚拟私有云”进入VPC服务查看已创建的VPC名称和ID。如果没有VPC，需要创建一个新的VPC。	vpc-default
子网	选择需要创建集群的子网，可进入VPC服务查看VPC下已创建的子网名称和ID。如果VPC下未创建子网，请单击“创建子网”进行创建。	subnet-default
集群节点	配置集群节点信息。	保持默认
Kerberos 认证	是否启用Kerberos认证。	不开启
用户名	登录集群管理页面及ECS节点的用户名。	admin/root
密码	设置登录集群管理页面及ECS节点用户的密码。	-
确认密码	再次输入设置用户密码。	-
企业项目	选择集群所属的企业项目。	default
通信安全授权	勾选确认授权。	勾选

图 2-1 购买实时分析集群



**步骤5** 单击“立即购买”，进入任务提交成功页面。

**步骤6** 单击“返回集群列表”，在“现有集群”列表中可以看到集群创建的状态。

集群创建需要时间，所创集群的初始状态为“启动中”，创建成功后状态更新为“运行中”，请您耐心等待。

----结束

## 步骤二：安装集群客户端

MRS集群创建成功后，用户可安装集群客户端用于连接集群内各组件服务，进行作业提交等操作。

客户端可以安装在集群内的节点上，也可以安装在集群外的节点上。本指南以在Master1节点上安装客户端为例进行介绍。

**步骤1** MRS集群创建成功后，在集群列表中单击MRS集群名称进入集群概览页面。

**步骤2** 单击“集群管理页面”后的“前往 Manager”，在弹出的窗口中选择“EIP访问”并配置弹性IP信息。

首次访问时，可点击“管理弹性公网IP”，在弹性公网IP控制台购买一个弹性公网IP，购买成功后刷新弹性公网IP列表并选择。

**步骤3** 勾选确认信息后，单击“确定”，登录集群的FusionInsight Manager管理界面。

Manager登录用户名为admin，密码为购买集群时配置的用于登录Manager管理界面的“admin”用户的自定义密码。

**步骤4** 在“主页”页签的集群名称后单击 **...**，单击“下载客户端”下载集群客户端。

图 2-2 下载客户端  
集群



在“下载集群客户端”弹窗中，配置以下参数：

- 选择客户端类型：选择“完整客户端”。
- 选择平台类型：默认勾选即可，例如“x86\_64”。
- 勾选“仅保存到如下路径”，使用默认保存路径，文件生成后将保存在集群主 OMS 节点的“/tmp/FusionInsight-Client”目录下。

图 2-3 下载集群客户端提示框

### 下载集群客户端

下载mrs\_ck的客户端，集群的客户端包括了所有服务

选择客户端类型： 完整客户端  仅配置文件

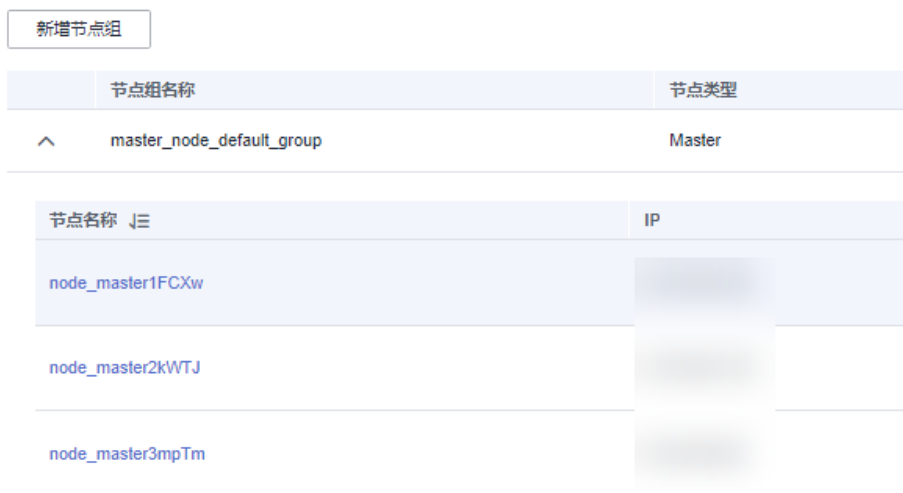
选择平台类型： x86\_64  aarch64

仅保存到如下路径： ?

单击“确定”后，等待客户端软件生成成功。

**步骤5** 在MRS服务管理控制台的集群列表中，单击集群名称，在集群的“节点管理”页签，单击包含“master1”的节点名称，在ECS详情页面单击右上角的“远程登录”登录Master1节点。

图 2-4 查看 Master1 节点



新增节点组	
节点组名称	节点类型
^ master_node_default_group	Master
节点名称	IP
node_master1FCXw	
node_master2kWTJ	
node_master3mpTm	

**步骤6** 使用root用户登录Master1节点，密码为购买集群时配置的用于登录集群节点的“root”用户自定义密码。

**步骤7** 执行以下命令切换到客户端软件包所在目录，并解压客户端软件包：

```
cd /tmp/FusionInsight-Client/
```

```
tar -xvf FusionInsight_Cluster_1_Services_Client.tar
```

```
tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar
```

**步骤8** 执行以下命令进入安装包所在目录，安装客户端：

```
cd FusionInsight_Cluster_1_Services_ClientConfig
```

执行如下命令安装客户端到指定目录（绝对路径），例如安装到“/opt/client”目录，等待客户端安装完成。

```
./install.sh /opt/client
```

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

#### 📖 说明

客户端安装目录可以不存在，会自动创建。但如果存在，则必须为空，目录路径不能包含空格。且客户端安装目录路径只能包含大写字母、小写字母、数字以及\_字符。

----结束

## 步骤三：使用 Kafka 客户端创建 Topic

**步骤1** 在集群列表中单击MRS集群名称进入集群概览页面。

**步骤2** 在“概览”页面单击“IAM用户同步”，勾选“全部同步”后单击“同步”，等待IAM用户同步完成。

**步骤3** 选择“组件管理 > ZooKeeper > 实例”，查看ZooKeeper quorumpeer角色实例的IP地址，记录任意一个IP地址。

图 2-5 查看 ZooKeeper 角色实例 IP



The screenshot shows the '实例' (Instances) tab for the '服务 ZooKeeper' (ZooKeeper Service). It contains a table with the following data:

角色	主机名	管理IP	业务IP
quorumpeer	node-master1cMln.mrs-1jgx.com		
quorumpeer	node-master2DMvU.mrs-1jgx.com		
quorumpeer	node-ana-coreliuBE.prs-1jgx.com		

**步骤4** 单击“服务配置”，查看ZooKeeper客户端连接端口“clientPort”参数值。

**步骤5** 单击“服务 ZooKeeper”返回组件列表。

图 2-6 返回组件列表



**步骤6** 选择“Kafka > 实例”，查看Kafka Broker角色实例的IP地址，记录任意一个IP地址。

图 2-7 查看 Broker 角色实例 IP



The screenshot shows the '实例' (Instances) tab for the '服务 Kafka' (Kafka Service). It contains a table with the following data:

角色	主机名	管理IP	业务IP
Broker	node-str-coresdHN		
Broker	node-str-coreBSVh		
Broker	node-str-coregYFm		

**步骤7** 单击“服务配置”，查看Kafka Broker连接端口“port”参数值。

**步骤8** 以root用户登录MRS客户端所在节点（Master1节点）。

**步骤9** 执行以下命令切换到客户端安装目录并配置环境变量。

```
cd /opt/client
```

```
source bigdata_env
```

**步骤10** 创建Kafka Topic。



```
kafka-topics.sh --create --zookeeper ZooKeeper角色实例IP地址:ZooKeeper客户端  
连接端口/kafka --partitions 2 --replication-factor 2 --topic Topic名称
```

例如执行以下命令：

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

执行结果输出以下信息表示创建Topic成功：

```
Created topic Topic1.
```

----结束

## 步骤四：管理 Kafka 主题中的消息

**步骤1** 以root用户登录MRS客户端所在节点（Master1节点）。

**步骤2** 执行以下命令切换到客户端安装目录并配置环境变量。

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

**步骤3** 在Topic1中产生消息。

```
kafka-console-producer.sh --broker-list Kafka Broker角色实例所在节点IP:Broker  
连接端口 --topic Topic名称 --producer.config /opt/hadoopclient/Kafka/kafka/  
config/producer.properties
```

Kafka Broker角色实例所在节点IP地址及端口信息可参考[步骤三：使用Kafka客户端创建Topic](#)的[步骤6](#)和[步骤7](#)获取。

例如执行以下命令：

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

**步骤4** 重新打开一个客户端连接窗口。

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

**步骤5** 执行命令消费Topic1中的消息。

```
kafka-console-consumer.sh --topic Topic名称 --bootstrap-server Kafka Broker角  
色实例所在节点IP:Broker连接端口 --consumer.config /opt/client/Kafka/kafka/  
config/consumer.properties
```

例如执行以下命令：

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

**步骤6** 在[步骤3](#)产生消息的命令中输入指定的内容作为生产者产生的消息，输入完成后按回车发送消息。

例如：

```
>aaa  
>bbb  
>ccc
```

如果需要结束产生消息，使用“Ctrl + C”退出任务。

**步骤7** 在**步骤5**的消费消息窗口中，可以观察到消息被成功消费。

```
aaa  
bbb  
ccc
```

----结束

## 相关信息

更多Kafka权限管理、Topic管理及消费、高可用配置及数据数据均衡等操作请参见[使用Kafka](#)。

# 3 快速创建和使用 HBase 离线数据查询集群

## 操作场景

本入门提供从零开始创建HBase查询集群并通过集群客户端进行HBase表的创建与查询操作指导。

HBase集群使用Hadoop和HBase组件提供一个稳定可靠、性能优异、可伸缩、面向列的分布式云存储系统，适用于海量数据存储以及分布式计算的场景，用户可以利用HBase搭建起TB至PB级数据规模的存储系统，对数据轻松进行过滤分析，毫秒级得到响应，快速发现数据价值。

## 操作流程

开始使用如下样例前，请务必按[准备工作](#)指导完成必要操作。

1. **步骤一：创建MRS集群**：自定义创建一个MRS 3.2.0-LTS.1版本的“HBase查询集群”。
2. **步骤二：安装集群客户端**：下载并安装MRS集群客户端。
3. **步骤三：使用HBase客户端创建表**：在HBase客户端创建表、插入表数据、查询表数据及删除表。

## 准备工作

- 注册账号并实名认证。  
在创建MRS集群之前，请先[注册华为账号并开通华为云](#)，进行[实名认证](#)。  
如果您已开通华为云并进行实名认证，请忽略此步骤。
- 已准备具有创建MRS集群的IAM用户，详细操作请参见[创建IAM用户并授权使用MRS](#)。

## 步骤一：创建 MRS 集群

**步骤1** 进入[购买MRS集群页面](#)。

**步骤2** 在服务列表中搜索“MapReduce服务 MRS”，进入MRS服务管理控制台。

**步骤3** 单击“购买集群”，进入“购买集群”页面，选择“快速购买”页签。

**步骤4** 根据实际业务规划情况填写集群配置信息（本示例为快速创建按需计费的MRS 3.2.0-LTS.1版本集群，如需了解更多参数配置请参考[快速创建MRS集群](#)）。

表 3-1 MRS 集群配置参数

参数名称	参数说明	取值样例
计费模式	选择待创建的集群的计费模式，MRS提供“包年/包月”与“按需计费”两种计费模式。 按需计费是一种后付费模式，即先使用再付费，按照MRS集群实际使用时长计费。	按需计费
区域	选择区域。 不同区域的云服务产品之间内网互不相通。请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。	中国-香港
集群名称	待创建的MRS集群名称。	mrs_demo
集群类型	根据实际业务需要选择待创建的MRS集群类型。“自定义”类型集群提供丰富的组件搭配，可自行选择对应版本MRS集群所支持的所有组件。	自定义
版本类型	待创建的MRS集群版本类型，不同版本所包含的开源组件版本及功能特性可能不同。	LTS版
集群版本	待创建的MRS集群版本，不同版本所包含的开源组件版本及功能特性可能不同，推荐选择最新版本。	MRS 3.2.0-LTS.1
组件选择	基于系统预置的集群模板选择要购买的集群组件。	HBase查询集群
可用区	选择集群工作区域下关联的可用区。	可用区1
虚拟私有云	选择需要创建集群的VPC，单击“查看虚拟私有云”进入VPC服务查看已创建的VPC名称和ID。如果没有VPC，需要创建一个新的VPC。	vpc-default
子网	选择需要创建集群的子网，可进入VPC服务查看VPC下已创建的子网名称和ID。如果VPC下未创建子网，请单击“创建子网”进行创建。	subnet-default
集群节点	配置集群节点信息。	保持默认
Kerberos 认证	是否启用Kerberos认证。	不开启
用户名	登录集群管理页面及ECS节点的用户名。	admin/root
密码	设置登录集群管理页面及ECS节点用户的密码。	-
确认密码	再次输入设置用户密码。	-
企业项目	选择集群所属的企业项目。	default
通信安全授权	勾选确认授权。	勾选

图 3-1 购买 HBase 查询集群

计费模式 ? 包年/包月 按需计费

集群名称

集群类型 自定义 更多类型

自定义集群

- 提供丰富的组件搭配，可自行选择对应版本MRS集群所支持的所有组件。
- 可按照业务需求自定义调整集群服务的部署方式，实现管理控制角色分开部署、管理控制角色合并部署或数据角色合并部署。
- 建议不要将多个数据存储类服务部署在一个节点组中，避免资源争抢。

版本类型 ? LTS版 普通版

集群版本

组件选择

实时分析集群	ClickHouse集群	Hadoop分析集群	HBase查询集群
Hadoop 3.3.1, Flink 1.15.0, Kafka 2.11-2.4.0, ZooKeeper 3.6.3, Ranger 2.0.0 and ClickHouse 22.3.2.2	ZooKeeper 3.6.3 and ClickHouse 22.3.2.2	Hadoop 3.3.1, Hive 3.1.0, Spark2x 3.1.1, Flink 1.15.0, ZooKeeper 3.6.3, Ranger 2.0.0 and Tez 0.9.2	Hadoop 3.3.1, HBase 2.2.3, ZooKeeper 3.6.3 and Ranger 2.0.0
海量数据采集，数据实时分析与查询	用于联机分析(OLAP)的列式数据库管理系统(DBMS)	海量数据分析与查询	海量数据存储，实现毫秒级数据查询

**步骤5** 单击“立即购买”，进入任务提交成功页面。

**步骤6** 单击“返回集群列表”，在“现有集群”列表中可以查看到集群创建的状态。

集群创建需要时间，所创集群的初始状态为“启动中”，创建成功后状态更新为“运行中”，请您耐心等待。

----结束

## 步骤二：安装集群客户端

MRS集群创建成功后，用户可安装集群客户端用于连接集群内各组件服务，进行作业提交等操作。

客户端可以安装在集群内的节点上，也可以安装在集群外的节点上。本指南以在Master1节点上安装客户端为例进行介绍。

**步骤1** MRS集群创建成功后，在集群列表中单击MRS集群名称进入集群概览页面。

**步骤2** 单击“集群管理页面”后的“前往 Manager”，在弹出的窗口中选择“EIP访问”并配置弹性IP信息。

首次访问时，可点击“管理弹性公网IP”，在弹性公网IP控制台购买一个弹性公网IP，购买成功后刷新弹性公网IP列表并选择。

**步骤3** 勾选确认信息后，单击“确定”，登录集群的FusionInsight Manager管理界面。

Manager登录用户名为admin，密码为购买集群时配置的用于登录Manager管理界面的“admin”用户的自定义密码。

**步骤4** 在“主页”页签的集群名称后单击 **...**，单击“下载客户端”下载集群客户端。

图 3-2 下载客户端  
集群



在“下载集群客户端”弹窗中，配置以下参数：

- 选择客户端类型：选择“完整客户端”。
- 选择平台类型：默认勾选即可，例如“x86\_64”。
- 勾选“仅保存到如下路径”，使用默认保存路径，文件生成后将保存在集群主 OMS 节点的“/tmp/FusionInsight-Client”目录下。

图 3-3 下载集群客户端提示框

### 下载集群客户端

下载mrs\_ck的客户端，集群的客户端包括了所有服务

选择客户端类型： 完整客户端  仅配置文件

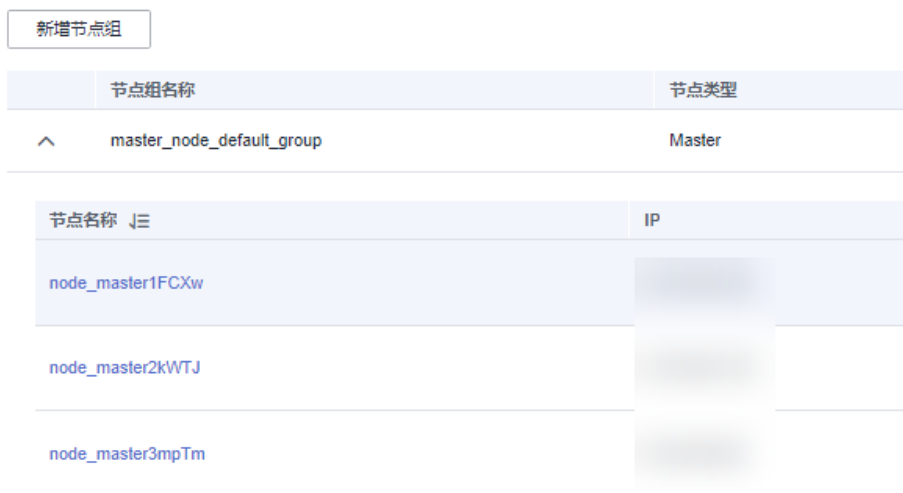
选择平台类型： x86\_64  aarch64

仅保存到如下路径：

单击“确定”后，等待客户端软件生成成功。

**步骤5** 在MRS服务管理控制台的集群列表中，单击集群名称，在集群的“节点管理”页签，单击包含“master1”的节点名称，在ECS详情页面单击右上角的“远程登录”登录Master1节点。

图 3-4 查看 Master1 节点



节点组名称	节点类型
^ master_node_default_group	Master

节点名称	IP
node_master1FCXw	
node_master2kWTJ	
node_master3mpTm	

**步骤6** 使用root用户登录Master1节点，密码为购买集群时配置的用于登录集群节点的“root”用户自定义密码。

**步骤7** 执行以下命令切换到客户端软件包所在目录，并解压客户端软件包：

```
cd /tmp/FusionInsight-Client/  
tar -xvf FusionInsight_Cluster_1_Services_Client.tar  
tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar
```

**步骤8** 执行以下命令进入安装包所在目录，安装客户端：

```
cd FusionInsight_Cluster_1_Services_ClientConfig
```

执行如下命令安装客户端到指定目录（绝对路径），例如安装到“/opt/client”目录，等待客户端安装完成。

```
./install.sh /opt/client
```

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

#### 📖 说明

客户端安装目录可以不存在，会自动创建。但如果存在，则必须为空，目录路径不能包含空格。且客户端安装目录路径只能包含大写字母、小写字母、数字以及\_字符。

----结束

## 步骤三：使用 HBase 客户端创建表

**步骤1** 以root用户登录MRS客户端所在节点（Master1节点）。

**步骤2** 执行以下命令切换到客户端安装目录并配置环境变量。

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

**步骤3** 执行以下命令进入HBase Shell命令行。

```
hbase shell
```

**步骤4** 运行HBase客户端命令，创建表“user\_info”并插入数据。

1. 执行以下命令创建表“user\_info”：

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

2. 执行以下命令向表“user\_info”中添加相关数据：

```
put 'user_info','12005000201','i:name','A'
put 'user_info','12005000201','i:gender','Male'
put 'user_info','12005000201','i:age','19'
put 'user_info','12005000201','i:address','City A'
put 'user_info','12005000201','i:degree','master'
put 'user_info','12005000201','i:pose','manager'
```

**步骤5** 查询HBase表内容。

1. 根据用户编号查询用户姓名和地址。

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

查询结果如下：

```
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
```

2. 根据用户姓名进行查询。

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

查询结果如下：

```
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
```

**步骤6** 删除HBase表。

1. 删除用户信息表中该用户的数据。

```
delete 'user_info','12005000201','i'
```

2. 删除用户信息表。

```
disable 'user_info'
drop 'user_info'
```

----结束



## 相关信息

更多HBase权限管理、使用索引及全局二级索引查询表数据、使用BulkLoad工具迁移HBase数据等操作请参见[使用HBase](#)。

# 4 快速创建和使用 ClickHouse 列式数据库集群

## 操作场景

本入门提供从零开始创建ClickHouse集群并通过集群客户端进行ClickHouse表的创建与查询操作指导。

ClickHouse是一款开源的面向联机分析处理的列式数据库，独立于Hadoop大数据体系，具有压缩率和极速查询性能。

## 操作流程

开始使用如下样例前，请务必按[准备工作](#)指导完成必要操作。

1. **步骤一：创建MRS集群**：自定义创建一个MRS 3.2.0-LTS.1版本的“ClickHouse集群”。
2. **步骤二：安装集群客户端**：下载并安装MRS集群客户端。
3. **步骤三：使用ClickHouse客户端创建表**：在ClickHouse客户端创建表并插入数据。

## 准备工作

- 注册账号并实名认证。  
在创建MRS集群之前，请先[注册华为账号并开通华为云](#)，进行**实名认证**。  
如果您已开通华为云并进行实名认证，请忽略此步骤。
- 已准备具有创建MRS集群的IAM用户，详细操作请参见[创建IAM用户并授权使用MRS](#)。

## 步骤一：创建 MRS 集群

**步骤1** 进入[购买MRS集群页面](#)。

**步骤2** 在服务列表中搜索“MapReduce服务 MRS”，进入MRS服务管理控制台。

**步骤3** 单击“购买集群”，进入“购买集群”页面，选择“快速购买”页签。

**步骤4** 根据实际业务规划情况填写集群配置信息（本示例为快速创建按需计费的MRS 3.2.0-LTS.1版本集群，如需了解更多参数配置请参考[快速创建MRS集群](#)）。

表 4-1 MRS 集群配置参数

参数名称	参数说明	取值
计费模式	选择待创建的集群的计费模式，MRS提供“包年/包月”与“按需计费”两种计费模式。 按需计费是一种后付费模式，即先使用再付费，按照MRS集群实际使用时长计费。	按需计费
区域	选择区域。 不同区域的云服务产品之间内网互不相通。请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。	中国-香港
集群名称	待创建的MRS集群名称。	mrs_demo
集群类型	根据业务实际需要选择待创建的MRS集群类型。“自定义”类型集群提供丰富的组件搭配，可自行选择对应版本MRS集群所支持的所有组件。	自定义
版本类型	待创建的MRS集群版本类型。	LTS版
集群版本	待创建的MRS集群版本，不同版本所包含的开源组件版本及功能特性可能不同，推荐选择最新版本。	MRS 3.2.0-LTS.1
组件选择	基于系统预置的集群模板选择要购买的集群组件。	ClickHouse集群
可用区	选择集群工作区域下关联的可用区。	可用区1
虚拟私有云	选择需要创建集群的VPC，单击“查看虚拟私有云”进入VPC服务查看已创建的VPC名称和ID。如果没有VPC，需要创建一个新的VPC。	vpc-default
子网	选择需要创建集群的子网，可进入VPC服务查看VPC下已创建的子网名称和ID。如果VPC下未创建子网，请单击“创建子网”进行创建。	subnet-default
集群节点	配置集群节点信息。	保持默认
Kerberos 认证	是否启用Kerberos认证。	不开启
用户名	登录集群管理页面及ECS节点的用户名。	admin/root
密码	设置登录集群管理页面及ECS节点用户的密码。	-
确认密码	再次输入设置用户密码。	-
企业项目	选择集群所属的企业项目。	default
通信安全授权	勾选确认授权。	勾选

图 4-1 购买 ClickHouse 集群

The screenshot shows the 'Purchase ClickHouse Cluster' configuration page in the MRS console. The cluster name is 'mrs\_demo'. The cluster type is 'Custom' (自定义), with a dropdown for 'More Types' (更多类型). Below this, there is a section for 'Custom Cluster' (自定义集群) with three bullet points: 'Provide rich component combinations, allowing selection of all components supported by the current MRS cluster version.', 'Adjust the deployment mode of the cluster services according to business requirements to realize the separation of management and control roles, or the combination of management and control roles, or the combination of management and control roles and data roles.', and 'Do not deploy multiple data storage services on a single node to avoid resource contention.' The version type is 'LTS' (LTS版) and the version is 'MRS 3.2.0-LTS.1'. There are four cluster options: 'Real-time Analysis Cluster' (实时分析集群) with Hadoop 3.3.1, Flink 1.15.0, Kafka 2.11-2.4.0, ZooKeeper 3.6.3, Ranger 2.0.0 and ClickHouse 22.3.2.2; 'ClickHouse Cluster' (ClickHouse集群) with ZooKeeper 3.6.3 and ClickHouse 22.3.2.2; 'Hadoop Analysis Cluster' (Hadoop分析集群) with Hadoop 3.3.1, Hive 3.1.0, Spark2x 3.1.1, Flink 1.15.0, ZooKeeper 3.6.3, Ranger 2.0.0 and Tez 0.9.2; and 'HBase Query Cluster' (HBase查询集群) with Hadoop 3.3.1, HBase 2.2.3, ZooKeeper 3.6.3 and Ranger 2.0.0. Each option has a brief description of its use case.

**步骤5** 单击“立即购买”，进入任务提交成功页面。

**步骤6** 单击“返回集群列表”，在“现有集群”列表中可以看到集群创建的状态。

集群创建需要时间，所创集群的初始状态为“启动中”，创建成功后状态更新为“运行中”，请您耐心等待。

----结束

## 步骤二：安装集群客户端

MRS集群创建成功后，用户可安装集群客户端用于连接集群内各组件服务，进行作业提交等操作。

客户端可以安装在集群内的节点上，也可以安装在集群外的节点上。本指南以在Master1节点上安装客户端为例进行介绍。

**步骤1** MRS集群创建成功后，在集群列表中单击MRS集群名称进入集群概览页面。

**步骤2** 单击“集群管理页面”后的“前往 Manager”，在弹出的窗口中选择“EIP访问”并配置弹性IP信息。

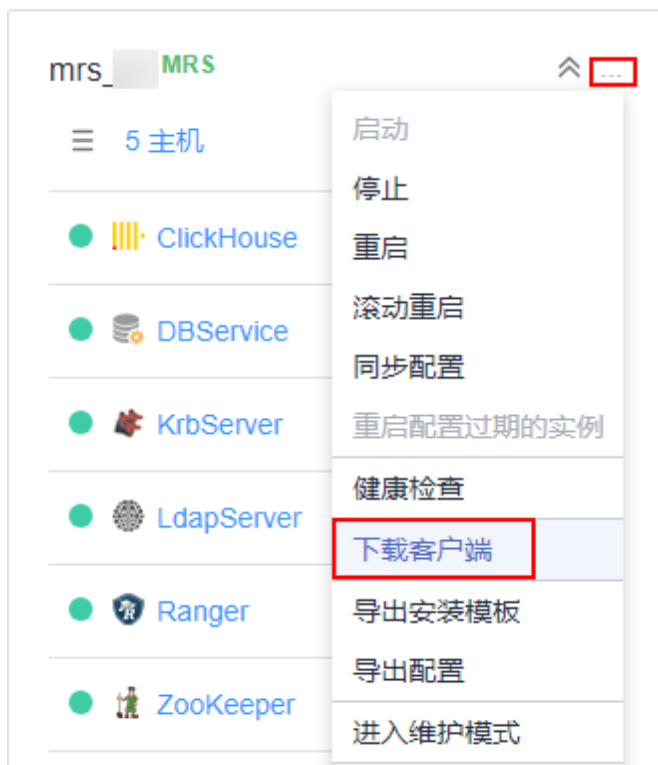
首次访问时，可点击“管理弹性公网IP”，在弹性公网IP控制台购买一个弹性公网IP，购买成功后刷新弹性公网IP列表并选择。

**步骤3** 勾选确认信息后，单击“确定”，登录集群的FusionInsight Manager管理界面。

Manager登录用户名为admin，密码为购买集群时配置的用于登录Manager管理界面的“admin”用户的自定义密码。

**步骤4** 在“主页”页签的集群名称后单击 **...**，单击“下载客户端”下载集群客户端。

图 4-2 下载客户端  
集群



在“下载集群客户端”弹窗中，配置以下参数：

- 选择客户端类型：选择“完整客户端”。
- 选择平台类型：默认勾选即可，例如“x86\_64”。
- 勾选“仅保存到如下路径”，使用默认保存路径，文件生成后将保存在集群主 OMS 节点的“/tmp/FusionInsight-Client”目录下。

图 4-3 下载集群客户端提示框

### 下载集群客户端

下载mrs\_ck的客户端，集群的客户端包括了所有服务

选择客户端类型： 完整客户端  仅配置文件

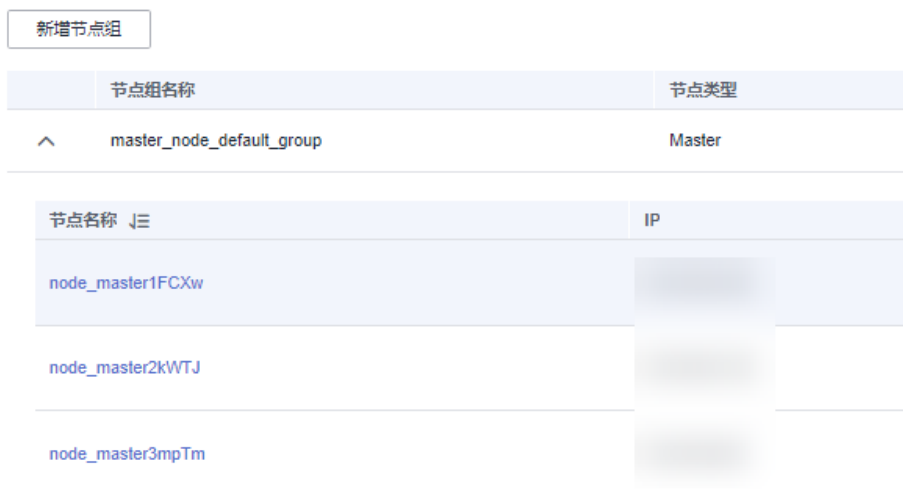
选择平台类型： x86\_64  aarch64

仅保存到如下路径：

单击“确定”后，等待客户端软件生成成功。

**步骤5** 在MRS服务管理控制台的集群列表中，单击集群名称，在集群的“节点管理”页签，单击包含“master1”的节点名称，在ECS详情页面单击右上角的“远程登录”登录Master1节点。

图 4-4 查看 Master1 节点



新增节点组	
节点组名称	节点类型
^ master_node_default_group	Master

节点名称	IP
node_master1FCXw	
node_master2kWTJ	
node_master3mpTm	

**步骤6** 使用root用户登录Master1节点，密码为购买集群时配置的用于登录集群节点的“root”用户自定义密码。

**步骤7** 执行以下命令切换到客户端软件包所在目录，并解压客户端软件包：

```
cd /tmp/FusionInsight-Client/  
tar -xvf FusionInsight_Cluster_1_Services_Client.tar  
tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar
```

**步骤8** 执行以下命令进入安装包所在目录，安装客户端：

```
cd FusionInsight_Cluster_1_Services_ClientConfig
```

执行如下命令安装客户端到指定目录（绝对路径），例如安装到“/opt/client”目录，等待客户端安装完成。

```
./install.sh /opt/client
```

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

#### 📖 说明

客户端安装目录可以不存在，会自动创建。但如果存在，则必须为空，目录路径不能包含空格。且客户端安装目录路径只能包含大写字母、小写字母、数字以及\_字符。

----结束

## 步骤三：使用 ClickHouse 客户端创建表

**步骤1** 以root用户登录MRS客户端所在节点（Master1节点）。

**步骤2** 执行以下命令切换到客户端安装目录并配置环境变量。

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

**步骤3** 使用clickhouse client命令连接ClickHouse服务端。

```
clickhouse client --host ClickHouseServer实例的IP地址 --port 9000 --user 用户名  
--password
```

#### 📖 说明

- ClickHouseServer实例的IP地址可登录集群FusionInsight Manager界面，选择“集群 > 服务 > ClickHouse > 实例”获取。
- 未开启Kerberos认证的集群默认使用非SSL安全连接，默认连接端口为**9000**，可登录集群FusionInsight Manager界面，选择“集群 > 服务 > ClickHouse > 配置”，搜索“tcp\_port”参数查看。
- 如果不指定--user和--password参数，默认使用default用户登录ClickHouse客户端。如果需要指定用户名和密码，则需在ClickHouse客户端执行create user SQL语句创建ClickHouse用户。

**步骤4** 执行以下命令创建数据库“test001”：

```
create database test001 on cluster default_cluster;
```

**步骤5** 执行以下命令创建复制表“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));
```

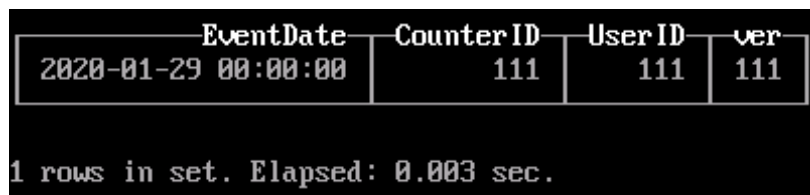
**步骤6** 执行以下命令创建复制表“test010”中插入数据：

```
insert into test001.test010 values('2020-01-29',111,111,111);
```

**步骤7** 执行以下命令查看“test010”表数据：

```
select * from test001.test010;
```

图 4-5 查看“test010”表数据



EventDate	Counter ID	User ID	ver
2020-01-29 00:00:00	111	111	111

1 rows in set. Elapsed: 0.003 sec.

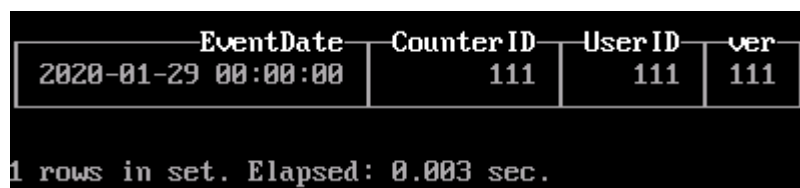
**步骤8** 执行以下命令创建分布式表“test001.test010\_dir”并插入数据：

```
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);
```

**步骤9** 执行以下命令查询分布式表：

```
select * from test001.test010_dir;
```

图 4-6 查看 “test010\_dir” 表数据



EventDate	Counter ID	User ID	ver
2020-01-29 00:00:00	111	111	111

1 rows in set. Elapsed: 0.003 sec.

**步骤10** 执行以下命令删除创建的库表：

```
drop database test001 on cluster default_cluster no delay;
```

----结束

## 相关信息

更多ClickHouse权限管理、导入RDS MySQL/OBS/HDFS/DWS等数据至ClickHouse表中、ClickHouse多租户管理、通过ELB访问ClickHouse等操作请参见[使用ClickHouse](#)。



# 5 快速创建和使用启用安全认证的 MRS 集群

## 操作场景

本入门提供从零开始创建开启Kerberos认证的Hadoop分析集群并通过集群客户端提交一个wordcount作业的操作指导。wordcount是最经典的Hadoop作业，用于统计海量文本的单词数量。

Hadoop集群完全使用开源Hadoop生态，采用Yarn管理集群资源，提供Hive、Spark离线大规模分布式数据存储和计算及进行海量数据分析与查询的能力。

## 操作流程

开始使用如下样例前，请务必按[准备工作](#)指导完成必要操作。

1. **步骤一：创建MRS集群**：创建一个MRS 3.2.0-LTS.1版本开启Kerberos认证的“Hadoop分析集群”。
2. **步骤二：创建集群用户**：在FusionInsight Manager界面创建一个具有提交wordcount作业权限的角色并绑定给用户。
3. **步骤三：安装集群客户端**：下载并安装MRS集群客户端。
4. **步骤四：准备应用程序及数据**：准备MRS集群客户端内wordcount样例程序运行所需的数据文件。
5. **步骤五：提交作业并查看结果**：在集群客户端提交wordcount数据分析作业并查看执行结果。

## 准备工作

- 注册账号并实名认证。  
在创建MRS集群之前，请先[注册华为账号并开通华为云](#)，进行[实名认证](#)。  
如果您已开通华为云并进行实名认证，请忽略此步骤。
- 已准备具有创建MRS集群的IAM用户，详细操作请参见[创建IAM用户并授权使用MRS](#)。

## 步骤一：创建 MRS 集群

**步骤1** 进入[购买MRS集群页面](#)。

**步骤2** 在服务列表中搜索“MapReduce服务 MRS”，进入MRS服务管理控制台。

**步骤3** 单击“购买集群”，进入“购买集群”页面，选择“快速购买”页签。

**步骤4** 根据实际业务规划情况填写集群配置信息（本示例为快速创建按需计费的MRS 3.2.0-LTS.1版本集群，如需了解更多参数配置请参考[快速创建MRS集群](#)）。

表 5-1 MRS 集群配置参数

参数名称	参数说明	取值样例
计费模式	选择待创建的集群的计费模式，MRS提供“包年/包月”与“按需计费”两种计费模式。 按需计费是一种后付费模式，即先使用再付费，按照MRS集群实际使用时长计费。	按需计费
区域	选择区域。 不同区域的云服务产品之间内网互不相通。请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。	中国-香港
集群名称	待创建的MRS集群名称。	mrs_demo
集群类型	根据业务实际需求选择待创建的MRS集群类型。“自定义”类型集群提供丰富的组件搭配，可自行选择对应版本MRS集群所支持的所有组件。	自定义
版本类型	待创建的MRS集群版本类型。	普通版
集群版本	待创建的MRS集群版本，不同版本所包含的开源组件版本及功能特性可能不同，推荐选择最新版本。	MRS 3.2.0-LTS.1
组件选择	基于系统预置的集群模板选择要购买的集群组件。	Hadoop分析集群
可用区	选择集群工作区域下关联的可用区。	可用区1
虚拟私有云	选择需要创建集群的VPC，单击“查看虚拟私有云”进入VPC服务查看已创建的VPC名称和ID。如果没有VPC，需要创建一个新的VPC。	vpc-default
子网	选择需要创建集群的子网，可进入VPC服务查看VPC下已创建的子网名称和ID。如果VPC下未创建子网，请单击“创建子网”进行创建。	subnet-default
集群节点	配置集群节点信息。	保持默认
Kerberos 认证	是否启用Kerberos认证。	开启
用户名	登录集群管理页面及ECS节点的用户名。	admin/root
密码	设置登录集群管理页面及ECS节点用户的密码。	-
确认密码	再次输入设置用户密码。	-
企业项目	选择集群所属的企业项目。	default

参数名称	参数说明	取值样例
通信安全授权	勾选确认授权。	勾选

图 5-1 购买 Hadoop 分析集群



**步骤5** 单击“立即购买”，进入任务提交成功页面。

**步骤6** 单击“返回集群列表”，在“现有集群”列表中可以看到集群创建的状态。

集群创建需要时间，所创集群的初始状态为“启动中”，创建成功后状态更新为“运行中”，请您耐心等待。

---结束

## 步骤二：创建集群用户

开启Kerberos认证的集群，必须通过以下步骤创建一个用户并分配相应权限来允许用户执行程序。

**步骤1** MRS集群创建成功后，在集群列表中单击MRS集群名称进入集群概览页面。

**步骤2** 单击“集群管理页面”后的“前往 Manager”，在弹出的窗口中选择“EIP访问”并配置弹性IP信息。

首次访问时，可点击“管理弹性公网IP”，在弹性公网IP控制台购买一个弹性公网IP，购买成功后刷新弹性公网IP列表并选择。

**步骤3** 勾选确认信息后，单击“确定”，登录集群的FusionInsight Manager管理界面。

Manager登录用户名为admin，密码为购买集群时配置的用于登录Manager管理界面的“admin”用户的自定义密码。

**步骤4** 选择“系统 > 权限 > 角色”。

**步骤5** 单击“添加角色”，配置如下参数创建一个角色，详情请参见[创建角色](#)。

- 填写角色的名称，例如mrrole。
- 在“配置资源权限”选择待操作的集群，然后选择“Yarn > 调度队列 > root”，勾选“权限”列中的“提交”和“管理”，勾选完全后，单击如下图的待操作的集群名，进行后面权限的选择。

图 5-2 配置 Yarn 服务权限

配置资源权限: 所有资源 > mr_role > Yarn > 调度队列 > root		权限	
资源名称	资源类型	提交	管理
launcher-job	子队列	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
default	子队列	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

- 选择“HDFS > 文件系统 > hdfs://hacluster/”，勾选“user”所在行的“权限”列的“读”、“写”和“执行”权限，单击“确定”，完成角色的创建。

图 5-3 配置 HDFS 服务权限

配置资源权限: 所有资源 > mr_role > HDFS > 文件系统		权限		
资源名称	资源类型	读	写	执行
hdfs://hacluster/	目录	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
viewfs://ClusterX/	目录	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**步骤6** 选择“用户 > 添加用户”，配置如下参数创建一个用户，详情请参见[创建用户](#)。

- 填写用户名，例如test。
- 用户类型为“人机”用户。
- 输入密码和确认新密码。
- 绑定“Manager\_viewer”和[步骤5](#)新建的“mrrole”角色取得权限。

图 5-4 添加用户

\* 用户名:  X

\* 用户类型:  人机  机机

\* 密码策略:

\* 密码:

\* 确认新密码:

用户组: [添加](#) | [清除全部](#) | [创建新用户组](#)

主组:

角色: [添加](#) | [清除全部](#) | [创建新角色](#)

mrrole XManager\_viewer X

**步骤7** 单击“确定”完成用户创建。

----结束

### 步骤三：安装集群客户端

MRS集群创建成功后，用户可安装集群客户端用于连接集群内各组件服务，进行作业提交等操作。

客户端可以安装在集群内的节点上，也可以安装在集群外的节点上。本指南以在 Master1 节点上安装客户端为例进行介绍。

**步骤1** MRS集群创建成功后，在集群列表中单击MRS集群名称进入集群概览页面。

**步骤2** 单击“集群管理页面”后的“前往 Manager”，在弹出的窗口中选择“EIP访问”并配置弹性IP信息。

首次访问时，可点击“管理弹性公网IP”，在弹性公网IP控制台购买一个弹性公网IP，购买成功后刷新弹性公网IP列表并选择。

**步骤3** 勾选确认信息后，单击“确定”，登录集群的FusionInsight Manager管理界面。

Manager登录用户名为admin，密码为购买集群时配置的用于登录Manager管理界面的“admin”用户的自定义密码。

**步骤4** 在“主页”页签的集群名称后单击...，单击“下载客户端”下载集群客户端。

图 5-5 下载客户端

### 集群



在“下载集群客户端”弹窗中，配置以下参数：

- 选择客户端类型：选择“完整客户端”。
- 选择平台类型：默认勾选即可，例如“x86\_64”。
- 勾选“仅保存到如下路径”，使用默认保存路径，文件生成后将保存在集群主 OMS 节点的“/tmp/FusionInsight-Client”目录下。

图 5-6 下载集群客户端提示框

## 下载集群客户端

下载mrs\_ck的客户端，集群的客户端包括了所有服务

选择客户端类型：

完整客户端

仅配置文件

选择平台类型：

x86\_64

aarch64

仅保存到如下路径：

/tmp/FusionInsight-Client/



确定

取消

单击“确定”后，等待客户端软件生成成功。

**步骤5** 在MRS服务管理控制台的集群列表中，单击集群名称，在集群的“节点管理”页签，单击包含“master1”的节点名称，在ECS详情页面单击右上角的“远程登录”登录Master1节点。

图 5-7 查看 Master1 节点

新增节点组

节点组名称	节点类型
^ master_node_default_group	Master

节点名称	IP
node_master1FCXw	
node_master2kWTJ	
node_master3mpTm	

**步骤6** 使用root用户登录Master1节点，密码为购买集群时配置的用于登录集群节点的“root”用户自定义密码。

**步骤7** 执行以下命令切换到客户端软件包所在目录，并解压客户端软件包：

```
cd /tmp/FusionInsight-Client/
```

```
tar -xvf FusionInsight_Cluster_1_Services_Client.tar
```

```
tar -xvf FusionInsight_Cluster_1_Services_ClientConfig.tar
```

**步骤8** 执行以下命令进入安装包所在目录，安装客户端：

```
cd FusionInsight_Cluster_1_Services_ClientConfig
```

执行如下命令安装客户端到指定目录（绝对路径），例如安装到“/opt/client”目录，等待客户端安装完成。

```
./install.sh /opt/client
```

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

#### 📖 说明

客户端安装目录可以不存在，会自动创建。但如果存在，则必须为空，目录路径不能包含空格。且客户端安装目录路径只能包含大写字母、小写字母、数字以及\_字符。

----结束

## 步骤四：准备应用程序及数据

MRS集群创建成功后，可获取集群客户端内的wordcount样例程序进行运行，也可准备上层业务自行开发的大数据应用程序并上传至集群。本指南以运行MRS集群客户端内的wordcount样例程序为例进行介绍，需先准备wordcount样例程序运行所需的数据文件。

**步骤1** 以root用户登录Master1节点。

**步骤2** 准备数据文件。

数据文件无格式要求，例如文件名为“wordcount1.txt”和“wordcount2.txt”，内容如下所示：

```
vi /opt/wordcount1.txt
```

```
hello word  
hello wordcount
```

```
vi /opt/wordcount2.txt
```

```
hello mapreduce  
hello hadoop
```

**步骤3** 执行以下命令切换到客户端安装目录，配置环境变量并创建用于存放样例数据的HDFS目录，例如“/user/example/input”。

```
cd /opt/client
```

```
source bigdata_env
```

```
kinit test ( test为步骤6创建的用户名，如果该用户首次登录，需修改初始密码)
```

```
hdfs dfs -mkdir /user/example
```

```
hdfs dfs -mkdir /user/example/input
```

#### 📖 说明

步骤6创建的test用户仅具有“/user”目录的“读”、“写”和“执行”权限，如果在“/user”之外其他目录中创建“input”目录会报权限不足错误，例如：

```
hdfs dfs -mkdir /hbase/input
```

命令执行后会报如下错误：

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



**步骤4** 执行以下命令，将样例数据上传至HDFS。

```
hdfs dfs -put /opt/wordcount1.txt /user/example/input
hdfs dfs -put /opt/wordcount2.txt /user/example/input
----结束
```

## 步骤五：提交作业并查看结果

**步骤1** 以root用户登录客户端节点（Master1节点）。

**步骤2** 执行如下命令提交wordcount作业，读取源数据进行分析并将执行结果输出至HDFS。

```
cd /opt/client
source bigdata_env
kinit test

hadoop jar HDFS/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-
examples-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
```

### 📖 说明

- “/user/example/output/”为HDFS上存放作业输出文件地址，请设置为一个不存在的目录。
- “hadoop-mapreduce-examples-3.3.1-\*.jar”文件在不同版本的集群客户端目录下文件名有差异，以实际名称为准。

**步骤3** 查看作业执行结果。

1. 作业执行完成后，可执行以下命令查看作业输出文件。

```
hdfs dfs -ls /user/example/output/
```

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

2. 输出结果保存在HDFS文件系统中，可以通过命令将其下载到本地进行查看。

例如执行以下命令：

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

“part-r-00000”文件内容如下：

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

**步骤4** 查看作业运行记录。

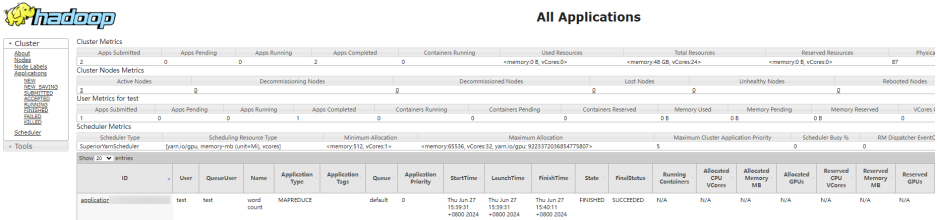
1. 使用**步骤6**创建的test用户登录集群的FusionInsight Manager页面，选择“集群 > 服务 > Yarn”。

2. 单击“ResourceManager Web UI”所在行的“ResourceManager(xxx, 主)”链接。
3. 在“All Applications”页面，单击目标作业的ID，可以查看作业运行的详情。

 说明

在“All Applications”页面可根据任务提交时间和提交任务的用户名确认具体任务。

图 5-8 查看作业详情



The screenshot shows the Hadoop 'All Applications' page. It includes a sidebar with navigation options like 'Cluster', 'Jobs', and 'Tools'. The main content area displays cluster metrics such as 'Apps Submitted', 'Apps Pending', 'Apps Running', etc. Below the metrics is a table of applications. The table has columns for ID, User, Owner/User, Name, Application Type, Application Type, Queue, Application Priority, StartTime, LaunchTime, FinishTime, State, FinalStatus, Running Containers, Allocated CPU VCore, Allocated Memory MB, Allocated GPU, Reserved CPU VCore, Reserved Memory MB, and Reserved GPU. One application is listed with ID 'application-', User 'test', Owner/User 'test', Name 'word-count', Application Type 'MAPREDUCE', Application Type 'default', Queue 'default', Application Priority '0', StartTime 'Thu Jun 27 15:39:31 +0800 2024', LaunchTime 'Thu Jun 27 15:39:31 +0800 2024', FinishTime 'Thu Jun 27 15:40:11 +0800 2024', State 'FINISHED', FinalStatus 'SUCCEEDED', and various resource allocation values.

----结束

### 相关信息

Hadoop组件包含了HDFS、Yarn、Mapreduce服务，可根据业务实际需求运行相关作业进行离线数据分析或查看，详细操作请参见[使用HDFS](#)、[使用MapReduce](#)或[使用Yarn](#)。

# 6 入门实践

当完成MRS集群部署后，可以根据自身的业务需求使用MRS提供的一系列常用实践。

表 6-1 MRS 常用最佳实践

实践	描述
数据分析	<b>使用Spark2x实现车联网车主驾驶行为分析</b> 本实践指导使用Spark实现车主驾驶行为分析。用于了解MRS的基本功能，利用MRS服务的Spark2x组件，对车主的驾驶行为进行分析统计，得到用户驾驶行为的分析结果，分析统计指定时间段内，车主急加速、急减速、空挡滑行、超速、疲劳驾驶等违法行为的次数。
	<b>使用Hive加载HDFS数据并分析图书评分情况</b> 本实践指导使用Hive对原始数据进行导入、分析等操作，展示了如何构建弹性、低成本的离线大数据分析。以某图书网站后台用户的点评数据为原始数据，导入Hive表后通过SQL命令筛选出最受欢迎的畅销图书。
	<b>使用Hive加载OBS数据并分析企业雇员信息</b> 本实践指导使用Hive对OBS中存储的原始数据进行导入、分析等操作，展示了如何构建弹性、低成本的存算分离大数据分析。以用户开发一个Hive数据分析应用为例，通过客户端连接Hive后，执行HQL语句访问OBS中的Hive数据。进行企业雇员信息的管理、查询。
	<b>通过Flink作业处理OBS数据</b> 本实践指导使用MRS集群内置的Flink WordCount作业程序，来分析OBS文件系统中保存的源数据，以统计源数据中的单词出现次数。 MRS支持在大数据存储容量大、计算资源需要弹性扩展的场景下，用户将数据存储在OBS服务中，使用MRS集群仅做数据计算处理的存算分离模式。
数据迁移	<b>数据迁移方案介绍</b> 本实践适用于多种不同场景下的HDFS、HBase、Hive数据向MRS集群的迁移工作。 介绍数据迁移前的准备工作、元数据导出、数据拷贝、数据恢复等内容。

实践	描述
<b>Hadoop数据迁移到华为云MRS服务</b>	本实践使用华为云CDM服务将Hadoop集群中的数据（支持数据量在几十TB级别或以下的数据量级）迁移到华为云MRS服务。
<b>HBase数据迁移到华为云MRS服务</b>	本实践使用华为云CDM服务将HBase集群中的数据（支持数据量在几十TB级别或以下的数据量级）迁移到华为云MRS服务。HBase会把数据存储到HDFS上，主要包括HFile文件和WAL文件，由配置项“hbase.rootdir”指定在HDFS上的路径，华为云MRS的默认存储位置是“/hbase”文件夹下。 HBase自带的一些机制和工具命令也可以实现数据搬迁，例如：通过导出Snapshots快照、Export/Import、CopyTable方式等。
<b>Hive数据迁移到华为云MRS服务</b>	本实践使用华为云CDM服务将Hive集群中的数据（支持数据量在几十TB级别或以下的数据量级）迁移到华为云MRS服务。 Hive数据迁移分两部分内容： <ul style="list-style-type: none"> <li>• Hive的元数据信息，存储在MySQL等数据库中。MRS Hive集群的元数据会默认存储到MRS DBService（华为的Gaussdb数据库），也可以选择RDS（MySQL）作为外置元数据库。</li> <li>• Hive的业务数据，存储在HDFS文件系统或OBS对象存储中。</li> </ul>
<b>MySQL数据迁移到MRS集群Hive分区表</b>	本实践使用CDM云服务将MySQL数据导入到MRS集群内的Hive分区表中。 Hive提供类SQL查询语言，帮助用户对大规模的数据进行提取、转换和加载，即通常所称的ETL（Extraction, Transformation, and Loading）操作。对庞大的数据集查询需要耗费大量的时间去处理，在许多场景下，可以通过建立Hive分区方法减少每一次扫描的总数据量，这种做法可以显著地改善性能。
<b>MRS HDFS数据迁移到OBS</b>	本实践以MRS HDFS数据迁移到OBS为例，介绍如何通过CDM将文件类数据迁移到文件中。
系统对接	
<b>使用DBEaver访问Phoenix</b>	本实践介绍如何使用DBEaver访问Phoenix。 本地安装的DBEaver可通过Phoenix Jar包对接MRS集群HBase组件。对接成功后，可在DBEaver中创建HBase表并插入数据。
<b>使用DBEaver访问HetuEngine</b>	本实践介绍如何使用DBEaver访问HetuEngine。 本地安装的DBEaver可通过JDBC Jar包对接MRS集群HetuEngine组件。对接成功后，可在DBEaver中查看HetuEngine已对接的数据源信息。

实践		描述
	<b>Hive对接外置自建关系型数据库</b>	<p>本实践介绍如何使用Hive对接开源MySQL和Postgres数据库。</p> <p>在已有Hive数据的集群上外置元数据库后，之前的元数据表不会自动同步。因此在安装Hive之初就要确认好元数据是外置数据库还是内置到DBService，如果是外置自建数据库，则需在安装Hive时或者暂无Hive数据时将元数据外置，安装后不允许修改，否则将会造成原有元数据丢失。</p>
	<b>Hive对接CSS服务</b>	<p>本实践介绍如何使用Hive对接CSS的Elasticsearch服务。</p> <p>利用Elasticsearch-Hadoop插件，完成Hive和CSS服务的Elasticsearch直接的数据交互，通过Hive外部表的方式，可以快速将Elasticsearch索引数据映射到Hive表中。</p>