



MapReduce 服务

快速入门

文档版本 01

发布日期 2019-06-05

华为技术有限公司



版权所有 © 华为技术有限公司 2019。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，华为公司对本文档内容不做任何明示或默示的声明或保证。

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。

目录

1 开启 Kerberos 认证集群..... 1

1 开启 Kerberos 认证集群

本章节提供从零开始使用安全集群并执行MapReduce程序、Spark程序和Hive程序的操作指导。

本指导的基本内容如下所示：

1. [创建安全集群并登录其Manager](#)
2. [创建角色和用户](#)
3. [执行MapReduce程序](#)
4. [执行Spark程序](#)
5. [执行Hive程序](#)

创建安全集群并登录其 Manager

当集群开启弹性公网IP功能时，执行如下步骤：

步骤1 创建安全集群，请参见[创建集群](#)页面，开启“Kerberos认证”参数开关，并配置“密码”、“确认密码”参数。该密码用于登录MRS Manager，请妥善保管。

图 1-1 安全集群参数配置

* Kerberos认证 ?

用户名 admin

该密码用于登录集群管理页面。

* 密码

* 确认密码

步骤2 登录MRS管理控制台页面。

步骤3 单击“集群列表”，在“现有集群”列表，单击指定的集群名称，进入集群信息页面。

步骤4 单击“点击查看”，打开“访问MRS Manager页面”。

- 若用户创建集群时已经绑定弹性公网IP，如图1-2所示。
 - a. 添加安全组规则，默认填充的是用户访问公网IP地址9022端口的规则。如需对安全组规则进行查看，修改和删除操作，请单击“管理安全组规则”。

 **说明**

- 自动获取的访问公网IP与用户本机IP不一致，属于正常现象，无需处理。
 - 9022端口为knox的端口，需要开启访问knox的9022端口权限，才能访问MRS Manager服务。
- b. 勾选“我确认xx.xx.xx.xx为可信任的公网访问IP，并允许从该IP访问MRS Manager页面。”

图 1-2 访问 MRS Manager 页面



- 若用户创建集群时暂未绑定弹性公网IP，如图1-3所示。
 - a. 在弹性公网IP下拉框中选择可用的弹性公网IP或单击“管理弹性公网IP”购买弹性公网IP。
 - b. 添加安全组规则，默认填充的是用户访问公网IP地址9022端口的规则。如需对安全组规则进行查看，修改和删除操作，请点击“管理安全组规则”。

 **说明**

- 自动获取的访问公网IP与用户本机IP不一致，属于正常现象，无需处理。
 - 9022端口为knox的端口，需要开启访问knox的9022端口权限，才能访问MRS Manager服务。
- c. 勾选“我确认xx.xx.xx.xx为可信任的公网访问IP，并允许从该IP访问MRS Manager页面。”

图 1-3 访问 MRS Manager 页面设置



步骤5 单击“确定”，进入MRS Manager登录页面，如需给其他用户开通访问MRS Manager的权限，请参见[访问支持Kerberos认证的Manager](#)章节，添加对应用户访问公网的IP地址为可信范围。

步骤6 输入创建集群时默认的用户名“admin”及设置的密码，单击“登录”进入MRS Manager页面。

----结束

当集群未开启弹性公网IP功能时，执行如下步骤：

步骤1 创建安全集群，请参见[创建集群](#)页面，开启“Kerberos认证”参数开关，并配置“密码”、“确认密码”参数。该密码用于登录MRS Manager，请妥善保管。

图 1-4 安全集群参数设置



步骤2 在MRS管理控制台，单击“集群列表”。

 说明

- 访问支持Kerberos认证的Manager方法请参考 [步骤2-步骤7](#)，或者请参考 [创建连接MRS集群的SSH隧道并配置浏览器](#)。
- 分析集群和流式集群访问支持Kerberos认证的Manager方法相同。

步骤3 在“现有集群”列表中，单击已创建的安全集群名称。

记录集群的“可用区”、“虚拟私有云”、“集群控制台地址”，以及Master节点的“默认安全组”。

步骤4 在弹性云服务器管理控制台，创建一个新的弹性云服务器。

- 弹性云服务器的“可用区”、“虚拟私有云”、“安全组”，需要和待访问集群的配置相同。
- 选择一个Windows系统的公共镜像。
- 其他配置参数详细信息，请参见[购买弹性云服务器](#)。

 说明

如果弹性云服务器的安全组和Master节点的“默认安全组”不同，用户可以选择以下任一种方法修改配置：

- 将弹性云服务器的安全组修改为Master节点的默认安全组，请参见[更改安全组](#)。
- 在集群Master节点和Core节点的安全组添加两条安全组规则使弹性云服务器可以访问集群，“协议”需选择为“TCP”，“端口”需分别选择“28443”和“20009”。请参见[创建安全组](#)。

步骤5 在虚拟私有云管理控制台，申请一个弹性IP地址，并与弹性云服务器绑定。

具体请参见[为弹性云服务器申请和绑定弹性公网IP](#)。

步骤6 登录弹性云服务器。

登录弹性云服务器需要Windows系统的帐号、密码，弹性IP地址以及配置安全组规则。具体请参见[Windows云服务器登录方式](#)。

步骤7 在Windows的远程桌面中，打开浏览器访问Manager。

例如Windows 2012操作系统可以使用Internet Explorer 11。

Manager访问地址形式为<https://集群控制台地址:28443/web>。**集群控制台地址**为**步骤3**中记录的“集群控制台地址”，访问时需要输入MRS集群的用户名和密码，例如“admin”用户，密码为**步骤1**创建集群时开启“Kerberos认证”参数开关时输入的密码。

 说明

- 如果使用其他MRS集群用户访问Manager，第一次访问时需要修改密码。新密码需要满足集群当前的用户密码复杂度策略。请咨询管理员。
- 默认情况下，在登录时输入5次错误密码将锁定用户，需等待5分钟自动解锁。

---结束

创建角色和用户

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 在MRS Manager界面选择“系统设置 > 角色管理”。

步骤2 单击“添加角色”，详情请参见[创建角色](#)。

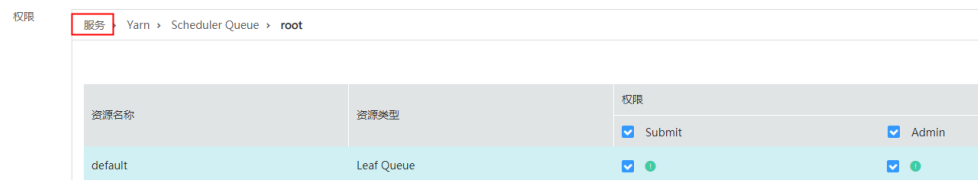
图 1-5 添加角色



填写如下信息：

- 填写角色的名称，例如mrrole。
- 在“权限”的表格中选择“Yarn > Scheduler Queue > root”，勾选“Submit”、“Admin”，勾选完全后，不要单击确认，要单击如下图的“服务”，再进行后面权限的选择。
- 在“权限”表格中选择“HBase > HBase Scope”，勾选global的“Create”、“Read”、“Write”、“Execute”，勾选完全后，不要单击确认，要单击如下图的“服务”，再进行后面权限的选择。
- 在“权限”的表格中选择“HDFS > File System > hdfs://hacluster/”，勾选“Read”、“Write”和“Execute”，勾选完全后，不要单击确认，要单击如下图的“服务”，再进行后面权限的选择。
- 在“权限”的表格中选择“Hive > Hive Read Write Privileges”，勾选default的“Create”、“Select”、“Delete”、“Insert”，单击“确定”，完成角色的创建。

图 1-6 配置服务权限



步骤3 单击“系统设置 > 用户组管理 > 添加用户组”，为样例工程创建一个用户组，例如mrgroup，详情请参见[创建用户组](#)。

步骤4 单击“系统设置 > 用户管理 > 添加用户”，为样例工程创建一个用户，详情请参见[创建用户](#)。

- 填写用户名，例如test，当需要执行Hive程序时，请设置用户名为“hiveuser”。
- 用户类型为“人机”用户。
- 输入密码（特别注意该密码在后面运行程序时要用到）。
- 加入用户组mrgroup和supergroup。
- 设置其“主组”为supergroup，并绑定角色mrrole取得权限。
单击“确定”完成用户创建。

步骤5 单击“系统设置 > 用户管理”，在用户名中选择新建用户“test”，单击“更多 > 下载认证凭据文件”，保存后解压得到用户的keytab文件与krb5.conf文件。

---结束

执行 MapReduce 程序

本小节提供执行MapReduce程序的操作指导，旨在指导用户在安全集群模式下运行程序。

前提条件

已编译好待运行的程序及对应的数据文件，如mapreduce-examples-1.0.jar、input_data1.txt和input_data2.txt，MapReduce程序开发及数据准备请参见[MapReduce应用开发](#)。

步骤1 采用远程登录软件（比如：MobaXterm）通过ssh登录（使用集群弹性IP登录）到安全集群的master节点。

步骤2 登录成功后分别执行下列命令，在/opt/client目录下创建test文件夹，在test目录下创建conf文件夹：

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

步骤3 使用上传工具（比如：WinScp）将mapreduce-examples-1.0.jar、input_data1.txt和input_data2.txt复制到test目录下，将“创建角色和用户”中的**步骤5**获得的keytab文件和krb5.conf文件复制到conf目录。

步骤4 执行如下命令配置环境变量并认证已创建用户，例如test。

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

然后按照提示输入密码，无异常提示返回，则完成了用户的kerberos认证。

步骤5 执行如下命令将数据导入到HDFS中：

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

步骤6 执行如下命令运行程序：

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

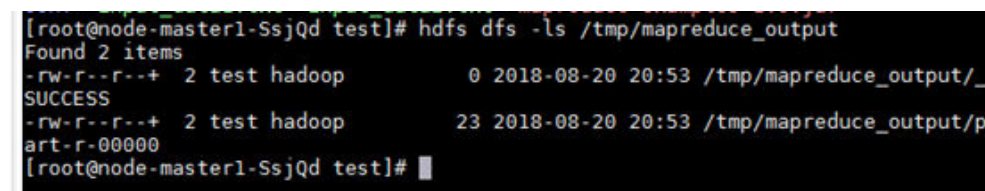
其中：

/tmp/input指HDFS文件系统中input的路径。

/tmp/mapreduce_output指HDFS文件系统中output的路径，该目录必须不存在，否则会报错。

步骤7 程序运行成功后，执行hdfs dfs -ls /tmp/mapreduce_output会显示如下：

图 1-7 查看程序运行结果



```
[root@node-master1-SsjQd test]# hdfs dfs -ls /tmp/mapreduce_output
Found 2 items
-rw-r--r--+ 2 test hadoop      0 2018-08-20 20:53 /tmp/mapreduce_output/_
SUCCESS
-rw-r--r--+ 2 test hadoop     23 2018-08-20 20:53 /tmp/mapreduce_output/p
art-r-00000
[root@node-master1-SsjQd test]#
```

---结束

执行 Spark 程序

本小节提供执行Spark程序的操作指导，旨在指导用户在安全集群模式下运行程序。

前提条件

已编译好待运行的程序及对应的数据文件，如FemaleInfoCollection.jar、input_data1.txt和input_data2.txt，Spark程序开发及数据准备请参见[Spark应用开发](#)。

步骤1 采用远程登录软件（比如：MobaXterm）通过ssh登录（使用集群弹性IP登录）到安全集群的master节点。

步骤2 登录成功后分别执行下列命令，在/opt/client目录下创建test文件夹，在test目录下创建conf文件夹：

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

步骤3 使用上传工具（比如：WinScp）将样FemaleInfoCollection.jar、input_data1.txt和input_data2.txt复制到test目录下，将“创建角色和用户”中的[步骤5](#)获得的keytab文件和krb5.conf文件复制到conf目录。

步骤4 执行如下命令配置环境变量并认证已创建用户，例如test。

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

然后按照提示输入密码，无异常提示返回，则完成了用户的kerberos认证。

步骤5 执行如下命令将数据导入到HDFS中：

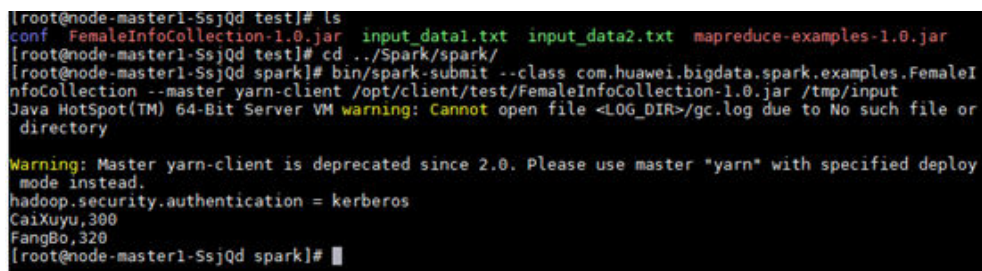
```
cd test
hdfs dfs -mkdir /tmp/input
hdfs dfs -put input_data* /tmp/input
```

步骤6 执行如下命令运行程序：

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

步骤7 程序运行成功后，会显示如下：

图 1-8 程序运行结果



```
[root@node-master1-SsjQd test]# ls
conf  FemaleInfoCollection-1.0.jar  input_data1.txt  input_data2.txt  mapreduce-examples-1.0.jar
[root@node-master1-SsjQd test]# cd ../Spark/spark/
[root@node-master1-SsjQd spark]# bin/spark-submit --class com.huawei.bigdata.spark.examples.FemaleI
nfoCollection --master yarn-client /opt/client/test/FemaleInfoCollection-1.0.jar /tmp/input
Java HotSpot(TM) 64-Bit Server VM warning: Cannot open file <LOG_DIR>/gc.log due to No such file or
directory
Warning: Master yarn-client is deprecated since 2.0. Please use master "yarn" with specified deploy
mode instead.
hadoop.security.authentication = kerberos
CaiXuyi,300
FangBo,320
[root@node-master1-SsjQd spark]#
```

----结束

执行 Hive 程序

本小节提供执行Hive程序的操作指导，旨在指导用户在安全集群模式下运行程序。

前提条件

已编译好待运行的程序及对应的数据文件，如hive-examples-1.0.jar、input_data1.txt和input_data2.txt，Hive程序开发及数据准备请参见[Hive应用开发](#)。

步骤1 采用远程登录软件（比如：MobaXterm）通过ssh登录（使用集群弹性IP登录）到安全集群的master节点。

步骤2 登录成功后分别执行下列命令，在/opt/client目录下创建test文件夹，在test目录下创建conf文件夹：

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

步骤3 使用上传工具（比如：WinScp）将样FemaleInfoCollection.jar、input_data1.txt和input_data2.txt复制到test目录下，将“创建角色和用户”中的[步骤5](#)获得的keytab文件和krb5.conf文件复制到conf目录。

步骤4 执行如下命令配置环境变量并认证已创建用户，例如test。

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

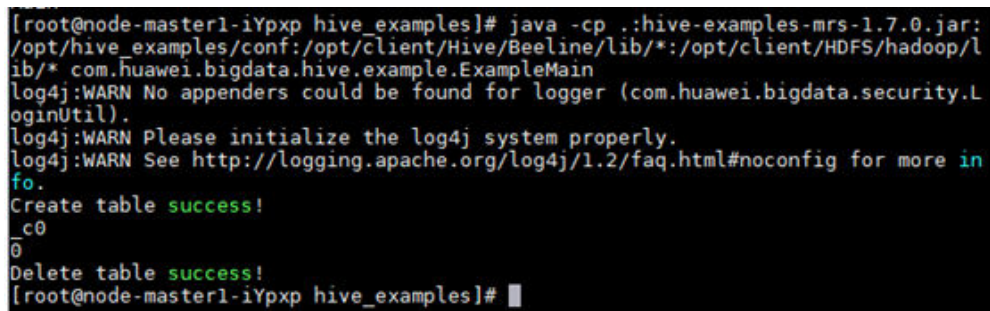
然后按照提示输入密码，无异常提示返回，则完成了用户的kerberos认证。

步骤5 执行如下命令运行程序：

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

步骤6 程序运行成功后，会显示如下：

图 1-9 程序运行的结果



```
[root@node-master1-iYpxp hive_examples]# java -cp ./hive-examples-mrs-1.7.0.jar:/opt/hive_examples/conf:/opt/client/Hive/Beeline/lib/*:/opt/client/HDFS/hadoop/lib/* com.huawei.bigdata.hive.example.ExampleMain
log4j:WARN No appenders could be found for logger (com.huawei.bigdata.security.LoginUtil).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
Create table success!
_c0
0
Delete table success!
[root@node-master1-iYpxp hive_examples]#
```

----结束