

Hadoop集成S3存储与hue配置

1.修改core-site.xml

修改master主机的core-site.xml

```
<!--配置s3 AccessKeyId-->
<property>
  <name>fs.s3a.access.key</name>
  <value>xxxxxxx</value> //填写自己的access.key
</property>
<!--配置s3 SecretAccessKey-->
<property>
  <name>fs.s3a.secret.key</name>
  <value>xxxxxxxxx/Q</value> //填写自己的secret.key
</property>
<!--配置s3 region-->
<property>
  <name>fs.s3a.endpoint</name>
  <value>s3.cn-northwest-1.amazonaws.com.cn</value> //可在aws官网查询桶所在的
endpoint
  <description>AWS S3 endpoint to connect to. An up-to-date list is
provided in the AWS Documentation: regions and endpoints. Without this
property, the standard region (s3.amazonaws.com) is assumed.
  </description>
</property>

<property>
  <name>fs.s3a.connection.ssl.enabled</name>
  <value>>false</value>
  <description>Enables or disables SSL connections to S3.</description>
</property>
```

将core-site.xml拷贝到其他slave主机（有文档说不需要拷贝也可以，未验证，建议拷贝）

2.导入S3相应的jar包

进入目录 /opt/dev/hadoop-2.8.5/share/hadoop/tools/lib

找到aws相关jar包，并复制到目录 /opt/dev/hadoop-2.8.5/share/hadoop/common/lib

```
[admin@sap-sjz1_os01 lib]$ ll |grep aws
-rw-r--r-- 1 admin admin 516062 Oct 9 15:09 aws-java-sdk-core-1.10.6.jar
-rw-r--r-- 1 admin admin 258578 Oct 9 15:09 aws-java-sdk-kms-1.10.6.jar
-rw-r--r-- 1 admin admin 570101 Oct 9 15:09 aws-java-sdk-s3-1.10.6.jar
-rw-r--r-- 1 admin admin 243779 Oct 9 15:09 hadoop-aws-2.8.5.jar
[admin@sap-sjz1_os01 lib]$ pwd
/opt/dev/hadoop-2.8.5/share/hadoop/tools/lib
[admin@sap-sjz1_os01 lib]$ cp aws*.jar /opt/dev/hadoop-
2.8.5/share/hadoop/common/lib
[admin@sap-sjz1_os01 lib]$ cp hadoop-aws-2.8.5.jar /opt/dev/hadoop-
2.8.5/share/hadoop/common/lib
```

找到joda-time-2.9.4.jar 并复制到目录: `/opt/dev/hadoop-2.8.5/share/hadoop/common/lib`

```
[admin@sap-sjzl_os01 lib]$ cp joda-time-2.9.4.jar /opt/dev/hadoop-2.8.5/share/hadoop/common/lib
```

找到jackson相关jar包并复制到目录: `/opt/dev/hadoop-2.8.5/share/hadoop/common/lib`

```
[admin@sap-sjzl_os01 lib]$ ll |grep jackson
-rw-r--r-- 1 admin admin 33483 Oct 11 16:09 jackson-annotations-2.2.3.jar
-rw-r--r-- 1 admin admin 192699 Oct 11 16:09 jackson-core-2.2.3.jar
-rw-r--r-- 1 admin admin 232248 Oct 11 16:09 jackson-core-asl-1.9.13.jar
-rw-r--r-- 1 admin admin 865838 Oct 11 16:09 jackson-databind-2.2.3.jar
-rw-r--r-- 1 admin admin 18336 Oct 11 16:09 jackson-jaxrs-1.9.13.jar
-rw-r--r-- 1 admin admin 780664 Oct 11 16:09 jackson-mapper-asl-1.9.13.jar
-rw-r--r-- 1 admin admin 27084 Oct 11 16:09 jackson-xc-1.9.13.jar
[admin@sap-sjzl_os01 lib]$ cp jackson*.jar /opt/dev/hadoop-2.8.5/share/hadoop/common/lib
```

!!! 注意每台主机均要移动jar包

3.重启hadoop并测试

重新启动hadoop

```
[admin@sap-sjzl_os01 lib]$ jps
47188 ResourceManager
49270 Jps
46999 SecondaryNameNode
46776 NameNode
4956 ThriftServer
1967 QuorumPeerMain
```

确认启动成功后, 输入神秘代码, 测试S3是否配置成功

```
[admin@sap-sjzl_os01 lib]$ hdfs dfs -ls s3a://hx-sjzl/
Found 1 items
-rw-rw-rw- 1 admin admin 284 2019-10-10 13:23 s3a://hx-sjzl/综合运维团队.txt
[admin@sap-sjzl_os01 lib]$
```

4.修改hue.ini配置

找到hue.ini配置文件 `/opt/dev/hue-4.2.0/desktop/conf`

vi 打开配置文件 使用 /s3 搜索到 aws 相关配置

```
#####
# Settings for the AWS lib
#####

[aws]
  [[aws_accounts]]
    # Default AWS account
    ## [[default]]
```

```
# AWS credentials
access_key_id=AKIAYZ6GC76KSFED6Z74
secret_access_key=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
## security_token=

# Execute this script to produce the AWS access key ID.
## access_key_id_script=/path/access_key_id.sh

# Execute this script to produce the AWS secret access key.
## secret_access_key_script=/path/secret_access_key.sh

# Allow to use either environment variables or
# EC2 InstanceProfile to retrieve AWS credentials.
## allow_environment_credentials=yes

# AWS region to use, if no region is specified, will attempt to connect to
standard s3.amazonaws.com endpoint
region=cn-northwest-1

# Endpoint overrides
host=s3.cn-northwest-1.amazonaws.com.cn
```

修改完成保存并重启。

5.hive导入S3相应的jar包

进入目录 `/opt/dev/hadoop-2.8.5/share/hadoop/tools/lib`

找到aws相关jar包，并复制到目录 `/opt/dev/hive-2.3.4/lib`

```
[admin@sap-sjz1_os01 lib]$ cp aws*.jar /opt/dev/hive-2.3.4/lib
[admin@sap-sjz1_os01 lib]$ cp hadoop-aws-2.8.5.jar /opt/dev/hive-2.3.4/lib
```

重启hive。

6.重启hue、hive并测试hive在S3上创建表

访问hue地址并进入sql编辑页面。输入建表语句，并测试

```
CREATE EXTERNAL TABLE s3_export(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3a://hx-sjz1/';
```

Hive

Add a name...

Add a descri...

test

text

1

2

3

CREATE EXTERNAL TABLE s3_export(a_col string, b_col bigint, c_col array<string>)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

LOCATION 's3a://hx-sjz1/';

▶

📖

Query History

🔍📅

Saved Queries

🔍

Query Builder

16 小时前

✓

CREATE EXTERNAL TABLE s3_export(a_col string, b_col

bigint, c_col array<string>)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

LOCATION 's3a://hx-sjz1/'

提示OK表示成功