

05-Hive

笔记本： 模块一配置笔记

创建时间： 2020/10/5 10:58

更新时间： 2021/7/18 22:58

作者： 封

五、Hive部署

0.软件包

涉及到的软件包，都存放在/opt/soft/

1.前置工作

请参照一前置工作操作

2.解压hive包

在master中，把Hive包解压到指定位置,按照要求是否需要重命名
tar -zxvf /opt/soft/apache-hive-2.3.7-bin.tar.gz -C /opt/soft/
重命名 mv /opt/soft/apache-hive-2.3.7 /opt/soft/hive

3.配置环境变量

在master中，vi /etc/profile

输入：

```
export HIVE_HOME=/opt/soft/hive
```

```
export PATH=$HIVE_HOME/bin:$PATH
```

```
source /etc/profile 更新环境变量
```

4.修改配置文件

在master中，复制hive-default.xml.template

复制：cp /opt /soft/hive/conf/hive-default.xml.template /opt/soft/conf/hive-default.xml 做为默认配置

新建或再复制hive-default.xml.template做为个性化配置

复制：cp /opt /soft/hive/conf/hive-default.xml.template /opt/soft/conf/hive-site.xml

新建并录入：vi /opt/soft/hive/conf/hive-site.xml

在hive-site.xml 文件中修改为以下配置：

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<?xml-stylesheet type="text/xsl" ?>
<configuration>
  <!-- ##### hive的 JDBC连接 ##### -->
  <!-- mysql 连接用户名 -->
  <property>
    <name>javax.jdo.option.ConnectionUserName</name>
    <value>root</value>
  </property>
  <!-- mysql 连接密码 -->
  <property>
    <name>javax.jdo.option.ConnectionPassword</name>
    <value>123456</value>
  </property>
  <!-- mysql 连接URL 如果hive和mysql在同一服务器上，使用localhost -->
  <property>
    <name>javax.jdo.option.ConnectionURL</name>
    <value>jdbc:mysql://master:3306/hive</value>
  </property>
  <!-- mysql 连接驱动 -->
  <property>
    <name>javax.jdo.option.ConnectionDriverName</name>
    <value>com.mysql.jdbc.Driver</value>
  </property>
</configuration>
```

5.上传驱动

在master中，利用XFtp把本地的mysql-connector-java-5.1.24.jar 上传到/opt/soft/hive/lib中即可。

6.mysql建库

在master中，登录mysql，新建数据库hive。或使用mysql的管理软件建立hive。

在master中，执行mysql -uroot -p，在提示中输入密码登录数据库，登录后，创建数据库。

指令：create database hive;

7.初始化Hive

在初始化前，需要把hadoop集群启动。

hadoop集群启动后，在master中，执行：schematool -dbType mysql -initSchema

在执行的完毕后，看到如下图显示，则已成功初始化。

```
Metastore connection URL: jdbc:mysql://cloud01:3306/hive?createDatabaseIfNotExist=true&characterEncoding=UTF-8
Metastore Connection Driver : com.mysql.jdbc.Driver
Metastore connection User: root
Starting metastore schema initialization to 2.1.0
Initialization script hive-schema-2.1.0.mysql.sql
Initialization script completed
schemaTool completed
```

8.进行Hive命令行客户端

在master中，执行hive。看到如下图显示，则已进入Hive命令行客户端。

```
Logging initialized using configuration in file:/home/dong/hive/conf/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. tez, spark) or using Hive 1.X releases.
hive> |
```

9.退出客户端

在hive客户端界面中，执行exit;

10.使用Hive

在hive的客户端界面中。

10.1显示所有hive数据库

输入：

```
show databases;
```

结果：

```
hive> show databases;  
OK  
default  
Time taken: 0.032 seconds, Fetched: 1 row(s)
```

10.2显示所有hive表

输入：

```
show tables;
```

结果：

```
hive> show tables;  
OK  
Time taken: 1.416 seconds
```

10.3建立一个books数据库

输入：

```
create database books;
```

结果：

```
hive> create database books;  
OK  
Time taken: 0.851 seconds
```

10.4查看数据库详细信息和文件系统位置

输入:

```
describe database books;
```

结果:

```
hive> describe database books;
OK
books          hdfs://cloud01:9000/user/dong/warehouse/books.db      dong
SER
Time taken: 0.042 seconds, Fetched: 1 row(s)
```

10.5在hdfs中查看刚刚建立的数据库books

在master中打开一个新的终端

输入:

```
hdfs dfs -ls /user/dong/warehouse
```

结果:

```
dong@cloud01:~$ hdfs dfs -ls /user/dong/warehouse
Found 1 items
drwxrwxrwx - dong supergroup          0 2018-03-11 16:41 /user/dong/warehouse/
books.db
```

10.6在books中创建内部表users（或受管理的表）

输入:

```
use books;
```

结果:

```
hive> use books;
OK
Time taken: 0.023 seconds
```

输入:

```
create table if not exists users(user_id INT,location STRING,age INT)
row format delimited fields terminated by '\;'      #设置表在加载数据
```

的时候，支持的列分隔符

```
stored as textfile; #设置加载数据的数据类型，如果文件数据是纯文本，就是使用 [STORED AS TEXTFILE]，然后从本地直接拷贝数据到HDFS上，hive直接可以识别数据
```

结果：

```
hive> create table if not exists users(user_id INT,location STRING,age INT)
> row format delimited fields terminated by '\;'
> stored as textfile;
OK
Time taken: 1.287 seconds
```

10.7查看表及表结构

输入：

```
show tables;
```

结果：

```
hive> show tables;
OK
users
Time taken: 0.125 seconds, Fetched: 1 row(s)
```

输入：

```
desc users; #或者describe users;
```

结果：

```
hive> desc users;
OK
user_id      int
location     string
age          int
Time taken: 0.515 seconds, Fetched: 3 row(s)
```

输入：

```
desc formatted users; #格式化查看表结构
```

```
OK
# col_name          data_type          comment
user_id             int
location            string
age                 int

# Detailed Table Information
Database:           books
Owner:              dong
CreateTime:         Sun Mar 11 17:29:44 CST 2018
LastAccessTime:     UNKNOWN
Retention:          0
Location:           hdfs://cloud01:9000/user/dong/warehouse/books.db/users
Table Type:         MANAGED_TABLE
Table Parameters:
  COLUMN_STATS_ACCURATE  {\ "BASIC_STATS\":\ "true\ "}
  numFiles               0
  numRows                0
  rawDataSize            0
  totalSize              0
  transient_lastDdlTime  1520760584

# Storage Information
SerDe Library:       org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat:         org.apache.hadoop.mapred.TextInputFormat
OutputFormat:        org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputForm
```

10.8查看数据

输入:

```
select * from users;
```

结果:

```
hive> select * from users;
OK
Time taken: 2.965 seconds
```

10.9使用load命令将数据加载到表。（该命令会将数据文件拷贝到HDFS中对应的位置）

输入:

```
load data local inpath "hive-data/Users-Prepro.txt" overwrite into table
users;
```

结果:

```
hive> load data local inpath 'hive-data/Users-Prepro.txt' overwrite into  
sers;  
Loading data to table books.users  
OK  
Time taken: 47.476 seconds
```

10.10查看表users的前十条数据:

输入:

```
select * from users limit 10;
```

结果:

```
hive> select * from users limit 10;  
OK  
1      nyc, new york, usa      NULL  
2      stockton, california, usa      18  
3      moscow, yukon territory, russia NULL  
4      porto, v.n.gaia, portugal      17  
5      farnborough, hants, united kingdom      NULL  
6      santa monica, california, usa      61  
7      washington, dc, usa      NULL  
8      timmins, ontario, canada      NULL  
9      germantown, tennessee, usa      NULL  
10     albacete, wisconsin, spain      26  
Time taken: 0.073 seconds, Fetched: 10 row(s)
```

10.11到hdfs中查看表数据

输入:

```
hdfs dfs -ls /user/dong/warehouse/books.db
```

结果:

```
dong@cloud01:~$ hdfs dfs -ls /user/dong/warehouse/books.db  
Found 1 items  
drwxrwxrwx - dong supergroup      0 2018-03-11 22:10 /user/dong/warehouse/  
books.db/users
```

输入:

```
hdfs dfs -ls /user/dong/warehouse/users/books.db/users
```


结果:

```
dong@cloud01:~$ hdfs dfs -ls /user/dong/warehouse/books.db/users
Found 1 items
-rwxrwxrwx   3 dong supergroup 10388093 2018-03-11 22:10 /user/dong/warehouse/
books.db/users/Users-Prepro.txt
```

10.12查询年龄在18~34岁之间的用户

输入:

```
select user_id,location,age from users where age>18 and age<34 limit 10;
```

结果:

```
hive> select user_id,location,age from users where age>18 and age<34 limit 10;
OK
10      albacete, wisconsin, spain      26
13      barcelona, barcelona, spain  26
18      rio de janeiro, rio de janeiro, brazil  25
20      langhorne, pennsylvania, usa  19
24      cologne, nrw, germany      19
27      chicago, illinois, usa     32
28      freiburg, baden-wuerttemberg, germany  24
29      cuernavaca, alabama, mexico  19
30      anchorage, alaska, usa      24
31      shanghai, n/a, china        20
Time taken: 1.091 seconds, Fetched: 10 row(s)
```

10.13统计年龄在18~34岁之间的用户总数

输入

```
select count(*) from users where age>18 and age<34;
```

结果:

```
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1520901017805_0001, Tracking URL = http://cloud01:8088/proxy/
application_1520901017805_0001/
Kill Command = /home/dong/hadoop-2.2.0/bin/hadoop job -kill job_1520901017805_0
001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-03-13 09:28:06,639 Stage-1 map = 0%, reduce = 0%
2018-03-13 09:28:13,226 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.88 se
c
2018-03-13 09:28:43,849 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.35
sec
MapReduce Total cumulative CPU time: 3 seconds 350 msec
Ended Job = job_1520901017805_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.35 sec HDFS Read: 1039696
9 HDFS Write: 105 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 350 msec
OK
74339
Time taken: 100.97 seconds, Fetched: 1 row(s)
```

10.14按年龄分组统计用户数量

输入:

```
select age, count(*) FROM users GROUP BY age;
```

结果:

OK		20	4047
NULL	110885	21	4426
0	416	22	4709
1	288	23	5450
2	105	24	5683
3	45	25	5614
4	28	26	5547
5	26	27	5373
6	18	28	5346
7	27	29	5289
8	54	30	4776
9	61	31	4663
10	84	32	4778
11	121	33	4697
12	192	34	4655
13	885	35	4194
14	1961	36	3896
15	2376	37	3582
16	2568	38	3403
17	3031	39	3234
18	3686	40	3142
19	3941	41	2919

10.15按年龄分组统计用户数量，并对结果按用户数量由高到低排序:

输入:

```
SELECT age, count(*) as c FROM users GROUP BY age ORDER BY c DESC;
```

结果:

OK	
NULL	110885
24	5683
25	5614
26	5547
23	5450
27	5373
28	5346
29	5289
32	4778
30	4776
22	4709
33	4697
31	4663
34	4655

10.15删除表users

输入:

```
drop table users;
```

结果:

```
hive> drop table users;  
OK  
Time taken: 4.239 seconds
```

10.16到hdfs中查看users是否还在

输入:

```
hdfs dfs -ls /user/dong/warehouse/books.db
```

结果:

```
dong@cloud01:~$ hdfs dfs -ls /user/dong/warehouse/books.db  
dong@cloud01:~$
```

10.17在hdfs的/user/dong/ 目录下创建books文件夹, 然后导入本地 (master主目录下) hive-data文件夹中的Books-Prepro.txt数据

输入:

```
hdfs dfs -mkdir /user/dong/books
```

结果:

```
dong@cloud01:~$ hdfs dfs -mkdir /user/dong/books  
dong@cloud01:~$ hdfs dfs -ls /user/dong/  
Found 2 items  
drwxr-xr-x - dong supergroup 0 2018-03-13 08:57 /user/dong/books  
drwxrwxrwx - dong supergroup 0 2018-03-11 16:41 /user/dong/warehouse
```

输入:

```
hdfs dfs -put hive-data/Books-Prepro.txt /user/dong/books
```

结果:

```
dong@cloud01:~$ hdfs dfs -ls /user/dong/books
Found 1 items
-rw-r--r--  3 dong supergroup    73402860 2018-03-13 08:59 /user/dong/books/Books-Prepro.txt
```

10.18创建外部数据表books

输入:

```
create external table if not exists books (
  isbn string,
  title string,
  author STRING,
  year INT,
  publisher STRING,
  image_s STRING,
  image_m STRING,
  image_l STRING)
row format delimited fields terminated by '\;'
stored as textfile
location '/user/dong/books'; #数据在hdfs上的位置
```

10.19查看外部表books的前3条数据

输入:

```
select * from books limit 3;
```

结果:

```
hive> select * from books limit 3;
OK
0195153448    Classical Mythology    Mark P. O. Morford    2002    Oxford Un
iversity Press  http://images.amazon.com/images/P/0195153448.01.THUMBZZZ.jpg  h
http://images.amazon.com/images/P/0195153448.01.MZZZZZZZ.jpg    http://images.ama
zon.com/images/P/0195153448.01.LZZZZZZZ.jpg
0002005018    Clara Callan    Richard Bruce Wright    2001    HarperFlamingo Ca
nada    http://images.amazon.com/images/P/0002005018.01.THUMBZZZ.jpg    http://im
ages.amazon.com/images/P/0002005018.01.MZZZZZZZ.jpg    http://images.amazon.com/
images/P/0002005018.01.LZZZZZZZ.jpg
0060973129    Decision in Normandy    Carlo D'Este    1991    HarperPerennial h
http://images.amazon.com/images/P/0060973129.01.THUMBZZZ.jpg    http://images.ama
zon.com/images/P/0060973129.01.MZZZZZZZ.jpg    http://images.amazon.com/images/P
/0060973129.01.LZZZZZZZ.jpg
Time taken: 0.143 seconds, Fetched: 3 row(s)
```

10.20删除外部表books

输入:

```
show tables;
```

结果:

```
hive> show tables;  
OK  
books
```

输入:

```
drop table books;
```

结果:

```
hive> drop table books;  
OK  
Time taken: 0.218 seconds
```

输入:

```
show tables;
```

结果:

```
hive> show tables;  
OK  
Time taken: 0.046 seconds
```

10.21到hdfs中查看books是否还在

输入:

```
hdfs dfs -ls /user/dong/
```

结果:

```
dong@cloud01:~$ hdfs dfs -ls /user/dong/  
Found 2 items  
drwxr-xr-x  - dong supergroup          0 2018-03-13 08:59 /user/dong/books  
drwxrwxrwx  - dong supergroup          0 2018-03-11 16:41 /user/dong/warehouse  
  
dong@cloud01:~$ hdfs dfs -ls /user/dong/books  
Found 1 items  
-rw-r--r--  3 dong supergroup    73402860 2018-03-13 08:59 /user/dong/books/Books-Prepro.txt
```

