

## 第 2 章 Hive 安装

### 2.1 Hive 安装地址

1) Hive 官网地址:

<http://hive.apache.org/>

2) 文档查看地址:

<https://cwiki.apache.org/confluence/display/Hive/GettingStarted>

3) 下载地址:

<http://archive.apache.org/dist/hive/>

4) github 地址:

<https://github.com/apache/hive>

### 2.2 Hive 安装部署

1) Hive 安装及配置

(1) 把 apache-hive-1.2.1-bin.tar.gz 上传到 linux 的/opt/software 目录下

(2) 解压 apache-hive-1.2.1-bin.tar.gz 到/opt/module/目录下

```
[atguigu@hadoop102 software]$ tar -zxvf apache-hive-1.2.1-bin.tar.gz -C /opt/module/
```

(3) 修改 apache-hive-1.2.1-bin.tar.gz 的名称为 hive

```
[atguigu@hadoop102 module]$ mv apache-hive-1.2.1-bin/ hive
```

(4) 修改/opt/module/hive/conf 目录下的 hive-env.sh.template 名称为 hive-env.sh

```
[atguigu@hadoop102 conf]$ mv hive-env.sh.template hive-env.sh
```

(5) 配置 hive-env.sh 文件

(a) 配置 HADOOP\_HOME 路径

```
export HADOOP_HOME=/opt/module/hadoop-2.7.2
```

(b) 配置 HIVE\_CONF\_DIR 路径

```
export HIVE_CONF_DIR=/opt/module/hive/conf
```

2) Hadoop 集群配置

(1) 必须启动 hdfs 和 yarn

```
[atguigu@hadoop102 hadoop-2.7.2]$ sbin/start-dfs.sh
```

```
[atguigu@hadoop103 hadoop-2.7.2]$ sbin/start-yarn.sh
```

- (2) 在 HDFS 上创建/tmp 和/user/hive/warehouse 两个目录并修改他们的同组权限可写

```
[atguigu@hadoop102 hadoop-2.7.2]$ bin/hadoop fs -mkdir /tmp
```

```
[atguigu@hadoop102 hadoop-2.7.2]$ bin/hadoop fs -mkdir -p /user/hive/warehouse
```

```
[atguigu@hadoop102 hadoop-2.7.2]$ bin/hadoop fs -chmod 777 /tmp
```

```
[atguigu@hadoop102 hadoop-2.7.2]$ bin/hadoop fs -chmod 777 /user/hive/warehouse
```

### 3) Hive 基本操作

- (1) 启动 hive

```
[atguigu@hadoop102 hive]$ bin/hive
```

- (2) 查看数据库

```
hive>show databases;
```

- (3) 打开默认数据库

```
hive>use default;
```

- (4) 显示 default 数据库中的表

```
hive>show tables;
```

- (5) 创建一张表

```
hive> create table student(id int, name string) ;
```

- (6) 显示数据库中有几张表

```
hive>show tables;
```

- (7) 查看表的结构

```
hive>desc student;
```

- (8) 向表中插入数据

```
hive> insert into student values(1000,"ss");
```

- (9) 查询表中数据

```
hive> select * from student;
```

- (10) 退出 hive

```
hive> quit;
```

## 2.3 将本地文件导入 Hive 案例

**需求:** 将本地/opt/module/datas/student.txt 这个目录下的数据导入到 hive 的 student(id int, name

string)表中。

1) 数据准备：在/opt/module/datas/student.txt 这个目录下准备数据

(1) 在/opt/module/目录下创建 datas

```
[atguigu@hadoop102 module]$ mkdir datas
```

(2) 在/opt/module/datas/目录下创建 student.txt 文件并添加数据

```
[atguigu@hadoop102 datas]$ touch student.txt
```

```
[atguigu@hadoop102 datas]$ vi student.txt
```

```
1001    zhangshan
```

```
1002    lishi
```

```
1003    zhaoliu
```

注意以 tab 键间隔。

2) Hive 实际操作

(1) 启动 hive

```
[atguigu@hadoop102 hive]$ bin/hive
```

(2) 显示数据库

```
hive>show databases;
```

(3) 使用 default 数据库

```
hive>use default;
```

(4) 显示 default 数据库中的表

```
hive>show tables;
```

(5) 删除已创建的 student 表

```
hive> drop table student;
```

(6) 创建 student 表，并声明文件分隔符'\t'

```
hive> create table student(id int, name string) ROW FORMAT DELIMITED FIELDS  
TERMINATED BY '\t';
```

(7) 加载/opt/module/datas/student.txt 文件到 student 数据库表中。

```
hive> load data local inpath '/opt/module/datas/student.txt' into table student;
```

(8) Hive 查询结果

```
hive> select * from student;
```

OK

1001      zhangshan

1002      lishi

1003      zhaoliu

Time taken: 0.266 seconds, Fetched: 3 row(s)

### 3) 遇到的问题

再打开一个客户端窗口启动 hive，会产生 java.sql.SQLException 异常。

```
Exception in thread "main" java.lang.RuntimeException: java.lang.RuntimeException: Unable
to instantiate org.apache.hadoop.hive.ql.metadata.SessionHiveMetaStoreClient
    at org.apache.hadoop.hive.ql.session.SessionState.start(SessionState.java:522)
    at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:677)
    at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:621)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at
sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
    at
sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:606)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
Caused by: java.lang.RuntimeException: Unable to instantiate
org.apache.hadoop.hive.ql.metadata.SessionHiveMetaStoreClient
    at
org.apache.hadoop.hive.metastore.MetaStoreUtils.newInstance(MetaStoreUtils.java:1523)
    at
org.apache.hadoop.hive.metastore.RetryingMetaStoreClient.<init>(RetryingMetaStoreClient.ja
va:86)
    at
org.apache.hadoop.hive.metastore.RetryingMetaStoreClient.getProxy(RetryingMetaStoreClien
t.java:132)
    at
org.apache.hadoop.hive.metastore.RetryingMetaStoreClient.getProxy(RetryingMetaStoreClien
t.java:104)
    at org.apache.hadoop.hive.ql.metadata.Hive.createMetaStoreClient(Hive.java:3005)
    at org.apache.hadoop.hive.ql.metadata.Hive.getMSC(Hive.java:3024)
    at org.apache.hadoop.hive.ql.session.SessionState.start(SessionState.java:503)
    ... 8 more
```

原因是，Metastore 默认存储在自带的 derby 数据库中，推荐使用 MySQL 存储 Metastore;

## 2.4 MySql 安装

### 2.4.1 安装包准备

1) 查看 mysql 是否安装，如果安装了，卸载 mysql

(1) 查看

```
[root@hadoop102 桌面]# rpm -qa|grep mysql
```

```
mysql-libs-5.1.73-7.el6.x86_64
```

(2) 卸载

```
[root@hadoop102 桌面]# rpm -e --nodeps mysql-libs-5.1.73-7.el6.x86_64
```

2) 解压 mysql-libs.zip 文件到当前目录

```
[root@hadoop102 software]# unzip mysql-libs.zip
```

```
[root@hadoop102 software]# ls
```

```
mysql-libs.zip
```

```
mysql-libs
```

3) 进入到 mysql-libs 文件夹下，并设置当前用户执行权限

```
[root@hadoop102 mysql-libs]# ll
```

```
总用量 76048
```

```
-rw-r--r--. 1 root root 18509960 3 月 26 2015 MySQL-client-5.6.24-1.el6.x86_64.rpm
```

```
-rw-r--r--. 1 root root 3575135 12 月 1 2013 mysql-connector-java-5.1.27.tar.gz
```

```
-rw-r--r--. 1 root root 55782196 3 月 26 2015 MySQL-server-5.6.24-1.el6.x86_64.rpm
```

```
[root@hadoop102 mysql-libs]# chmod u+x ./*
```

```
[root@hadoop102 mysql-libs]# ll
```

```
总用量 76048
```

```
-rwxr--r--. 1 root root 18509960 3 月 26 2015 MySQL-client-5.6.24-1.el6.x86_64.rpm
```

```
-rwxr--r--. 1 root root 3575135 12 月 1 2013 mysql-connector-java-5.1.27.tar.gz
```

```
-rwxr--r--. 1 root root 55782196 3 月 26 2015 MySQL-server-5.6.24-1.el6.x86_64.rpm
```

### 2.4.2 安装 MySql 服务器

1) 安装 mysql 服务端

```
[root@hadoop102 mysql-libs]# rpm -ivh MySQL-server-5.6.24-1.el6.x86_64.rpm
```

- 2) 查看产生的随机密码

```
[root@hadoop102 mysql-libs]# cat /root/.mysql_secret
```

```
OEXaQuS8IWkG19Xs
```

- 3) 查看 mysql 状态

```
[root@hadoop102 mysql-libs]# service mysql status
```

- 4) 启动 mysql

```
[root@hadoop102 mysql-libs]# service mysql start
```

### 2.4.3 安装 MySQL 客户端

- 1) 安装 mysql 客户端

```
[root@hadoop102 mysql-libs]# rpm -ivh MySQL-client-5.6.24-1.el6.x86_64.rpm
```

- 2) 链接 mysql

```
[root@hadoop102 mysql-libs]# mysql -uroot -pOEXaQuS8IWkG19Xs
```

- 3) 修改密码

```
mysql>SET PASSWORD=PASSWORD('000000');
```

- 4) 退出 mysql

```
mysql>exit
```

### 2.4.4 MySQL 中 user 表中主机配置

配置只要是 root 用户+密码，在任何主机上都能登录 MySQL 数据库。

- 1) 进入 mysql

```
[root@hadoop102 mysql-libs]# mysql -uroot -p000000
```

- 2) 显示数据库

```
mysql>show databases;
```

- 3) 使用 mysql 数据库

```
mysql>use mysql;
```

- 4) 展示 mysql 数据库中的所有表

```
mysql>show tables;
```

- 5) 展示 user 表的结构

```
mysql>desc user;
```

6) 查询 user 表

```
mysql>select User, Host, Password from user;
```

7) 修改 user 表, 把 Host 表内容修改为%

```
mysql>update user set host='%' where host='localhost';
```

8) 删除 root 用户的其他 host

```
mysql>delete from user where Host='hadoop102 ';
```

```
mysql>delete from user where Host='127.0.0.1';
```

```
mysql>delete from user where Host='::1';
```

9) 刷新

```
mysql>flush privileges;
```

10) 退出

```
mysql> quit;
```

## 2.5 Hive 元数据配置到 MySQL

### 2.5.1 驱动拷贝

1) 在/opt/software/mysql-lib 目录下解压 mysql-connector-java-5.1.27.tar.gz 驱动包

```
[root@hadoop102 mysql-lib]# tar -zxvf mysql-connector-java-5.1.27.tar.gz
```

2) 拷贝/opt/software/mysql-lib/mysql-connector-java-5.1.27 目录下的

mysql-connector-java-5.1.27-bin.jar 到/opt/module/hive/lib/

```
[root@hadoop102 mysql-connector-java-5.1.27]# cp mysql-connector-java-5.1.27-bin.jar  
/opt/module/hive/lib/
```

### 2.5.2 配置 Metastore 到 MySQL

1) 在/opt/module/hive/conf 目录下创建一个 hive-site.xml

```
[atguigu@hadoop102 conf]$ touch hive-site.xml
```

```
[atguigu@hadoop102 conf]$ vi hive-site.xml
```

2) 根据官方文档配置参数, 拷贝数据到 hive-site.xml 文件中。

<https://cwiki.apache.org/confluence/display/Hive/AdminManual+MetastoreAdmin>

```
<?xml version="1.0"?>  
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>  
<configuration>
```

```
<property>
  <name>javax.jdo.option.ConnectionURL</name>

<value>jdbc:mysql://hadoop102:3306/metastore?createDatabaseIfNotExist=true</value>
  <description>JDBC connect string for a JDBC metastore</description>
</property>

<property>
  <name>javax.jdo.option.ConnectionDriverName</name>
  <value>com.mysql.jdbc.Driver</value>
  <description>Driver class name for a JDBC metastore</description>
</property>

<property>
  <name>javax.jdo.option.ConnectionUserName</name>
  <value>root</value>
  <description>username to use against metastore database</description>
</property>

<property>
  <name>javax.jdo.option.ConnectionPassword</name>
  <value>000000</value>
  <description>password to use against metastore database</description>
</property>
</configuration>
```

3) 配置完毕后, 如果启动 hive 异常, 可以重新启动虚拟机。(重启后, 别忘了启动 hadoop 集群)

### 2.5.3 多窗口启动 Hive 测试

1) 先启动 MySQL

```
[atguigu@hadoop102 mysql-lib]$ mysql -uroot -p000000
```

查看有几个数据库

```
mysql> show databases;
```

```
+-----+
| Database          |
+-----+
| information_schema |
| mysql              |
| performance_schema |
```



```
| test |
+-----+
```

- 2) 再次打开多个窗口，分别启动 hive

```
[atguigu@hadoop102 hive]$ bin/hive
```

- 3) 启动 hive 后，回到 MySQL 窗口查看数据库，显示增加了 metastore 数据库

```
mysql> show databases;
```

```
+-----+
| Database |
+-----+
| information_schema |
| metastore |
| mysql |
| performance_schema |
| test |
+-----+
```

## 2.6 Hive 常用交互命令

```
[atguigu@hadoop102 hive]$ bin/hive -help
```

usage: hive

-d,--define <key=value>	Variable substitution to apply to hive commands. e.g. -d A=B or --define A=B
--database <databasename>	Specify the database to use
-e <quoted-query-string>	SQL from command line
-f <filename>	SQL from files
-H,--help	Print help information
--hiveconf <property=value>	Use value for given property
--hivevar <key=value>	Variable substitution to apply to hive commands. e.g. --hivevar A=B
-i <filename>	Initialization SQL file
-S,--silent	Silent mode in interactive shell
-v,--verbose	Verbose mode (echo executed SQL to the console)

- 1) “-e”不进入 hive 的交互窗口执行 sql 语句

```
[atguigu@hadoop102 hive]$ bin/hive -e "select id from student;"
```

- 2) “-f”执行脚本中 sql 语句

- (1) 在/opt/module/datas 目录下创建 hivef.sql 文件

```
[atguigu@hadoop102 datas]$ touch hivef.sql
```

文件中写入正确的 sql 语句

```
select *from student;
```

- (2) 执行文件中的 sql 语句

```
[atguigu@hadoop102 hive]$ bin/hive -f /opt/module/datas/hivef.sql
```

- (3) 执行文件中的 sql 语句并将结果写入文件中

```
[atguigu@hadoop102 hive]$ bin/hive -f /opt/module/datas/hivef.sql >
/opt/module/datas/hive_result.txt
```

## 2.7 Hive 其他命令操作

- 1) 退出 hive 窗口:

```
hive(default)>exit;
```

```
hive(default)>quit;
```

在新版的 oracle 中没区别了, 在以前的版本是有的:

exit:先隐性提交数据, 再退出;

quit:不提交数据, 退出;

- 2) 在 hive cli 命令窗口中如何查看 hdfs 文件系统

```
hive(default)>dfs -ls /;
```

- 3) 在 hive cli 命令窗口中如何查看 hdfs 本地系统

```
hive(default)>! ls /opt/module/datas;
```

- 4) 查看在 hive 中输入的所有历史命令

- (1) 进入到当前用户的根目录/root 或/home/atguigu

- (2) 查看 .hivehistory 文件

```
[atguigu@hadoop102 ~]$ cat .hivehistory
```

## 2.8 Hive 常见属性配置

### 2.8.1 Hive 数据仓库位置配置

- 1) Default 数据仓库的最原始位置是在 hdfs 上的: /user/hive/warehouse 路径下

- 2) 在仓库目录下, 没有对默认的数据库 default 创建文件夹。如果某张表属于 default

数据库，直接在数据仓库目录下创建一个文件夹。

3) 修改 default 数据仓库原始位置 (将 hive-default.xml.template 如下配置信息拷贝到 hive-site.xml 文件中)

```
<property>
  <name>hive.metastore.warehouse.dir</name>
  <value>/user/hive/warehouse</value>
  <description>location of default database for the warehouse</description>
</property>
```

配置同组用户有执行权限

```
bin/hdfs dfs -chmod g+w /user/hive/warehouse
```

## 2.8.2 查询后信息显示配置

1) 在 hive-site.xml 文件中添加如下配置信息，就可以实现显示当前数据库，以及查询表的头信息配置。

```
<property>
  <name>hive.cli.print.header</name>
  <value>true</value>
</property>

<property>
  <name>hive.cli.print.current.db</name>
  <value>true</value>
</property>
```

2) 重新启动 hive，对比配置前后差异

(1) 配置前

```
hive> select * from student;
OK
1001    xiaoli
1002    libingbing
1003    fanbingbing
Time taken: 0.318 seconds, Fetched: 3 row(s)
```

(2) 配置后

```
hive (db_hive)> select * from db_hive.student;
OK
student.id    student.name
1001          xiaoli
1002          libingbing
1003          fanbingbing
```

### 2.8.3 Hive 运行日志信息配置

1) Hive 的 log 默认存放在/tmp/atguigu/hive.log 目录下（当前用户名下）。

2) 修改 hive 的 log 存放日志到/opt/module/hive/logs

（1）修改/opt/module/hive/conf/hive-log4j.properties.template 文件名称为

hive-log4j.properties

```
[atguigu@hadoop102 conf]$ pwd
```

```
/opt/module/hive/conf
```

```
[atguigu@hadoop102 conf]$ mv hive-log4j.properties.template hive-log4j.properties
```

（2）在 hive-log4j.properties 文件中修改 log 存放位置

```
hive.log.dir=/opt/module/hive/logs
```

### 2.8.4 参数配置方式

1) 查看当前所有的配置信息

```
hive>set;
```

2) 参数的配置三种方式

（1）配置文件方式

默认配置文件：hive-default.xml

用户自定义配置文件：hive-site.xml

注意：用户自定义配置会覆盖默认配置。另外，Hive 也会读入 Hadoop 的配置，因为 Hive 是作为 Hadoop 的客户端启动的，Hive 的配置会覆盖 Hadoop 的配置。配置文件的设定对本机启动的所有 Hive 进程都有效。

（2）命令行参数方式

启动 Hive 时，可以在命令行添加-hiveconf param=value 来设定参数。

例如：

```
[atguigu@hadoop103 hive]$ bin/hive -hiveconf mapred.reduce.tasks=10;
```

注意：仅对本次 hive 启动有效

查看参数设置：

```
hive (default)> set mapred.reduce.tasks;
```

（3）参数声明方式

可以在 HQL 中使用 SET 关键字设定参数

例如：

```
hive (default)> set mapred.reduce.tasks=100;
```

注意：仅对本次 hive 启动有效。

查看参数设置

```
hive (default)> set mapred.reduce.tasks;
```

上述三种设定方式的优先级依次递增。即配置文件<命令行参数<参数声明。注意某些系统级的参数，例如 log4j 相关的设定，必须用前两种方式设定，因为那些参数的读取在会话建立以前已经完成了。