

Sqoop 介绍、安装与操作

本文版权归作者和博客园共有，欢迎转载，但未经作者同意必须保留此段声明，且在文章页面明显位置给出原文连接，博主为石山园，博客地址为 <http://www.cnblogs.com/shishanyuan> 。该系列课程是应邀实验楼整理编写的，这里需要赞一下实验楼提供了学习的新方式，可以边看博客边上机实验，课程地址为 <https://www.shiyanlou.com/courses/237>

【注】该系列所使用到安装包、测试数据和代码均可在百度网盘下载，具体地址为 <http://pan.baidu.com/s/10PnDs>，下载该 PDF 文件

1 搭建环境

部署节点操作系统为 CentOS，防火墙和 SELinux 禁用，创建了一个 shiyanlou 用户并在系统根目录下创建/app 目录，用于存放 Hadoop 等组件运行包。因为该目录用于安装 hadoop 等组件程序，用户对 shiyanlou 必须赋予 rwx 权限（一般做法是 root 用户在根目录下创建/app 目录，并修改该目录拥有者为 shiyanlou(chown -R shiyanlou:shiyanlou /app)）。

Hadoop 搭建环境：

- 虚拟机操作系统：CentOS6.6 64 位，单核，1G 内存
- JDK：1.7.0_55 64 位
- Hadoop：1.1.2

2 Sqoop 介绍

2.1 Sqoop 简介

Sqoop 即 SQL to Hadoop，是一款方便的在传统型数据库与 Hadoop 之间进行数据迁移的工具，充分利用 MapReduce 并行特点以批处理的方式加快数据传输，发展至今主要演化了两大版本，Sqoop1 和 Sqoop2。

Sqoop 工具是 hadoop 下连接关系型数据库和 Hadoop 的桥梁，支持关系型数据库和 hive、hdfs，hbase 之间数据的相互导入，可以使用全表导入和增量导入。

那么为什么选择 Sqoop 呢？

- 高效可控的利用资源，任务并行度，超时时间。
- 数据类型映射与转化，可自动进行，用户也可自定义

- 支持多种主流数据库，MySQL, Oracle，SQL Server，DB2 等等

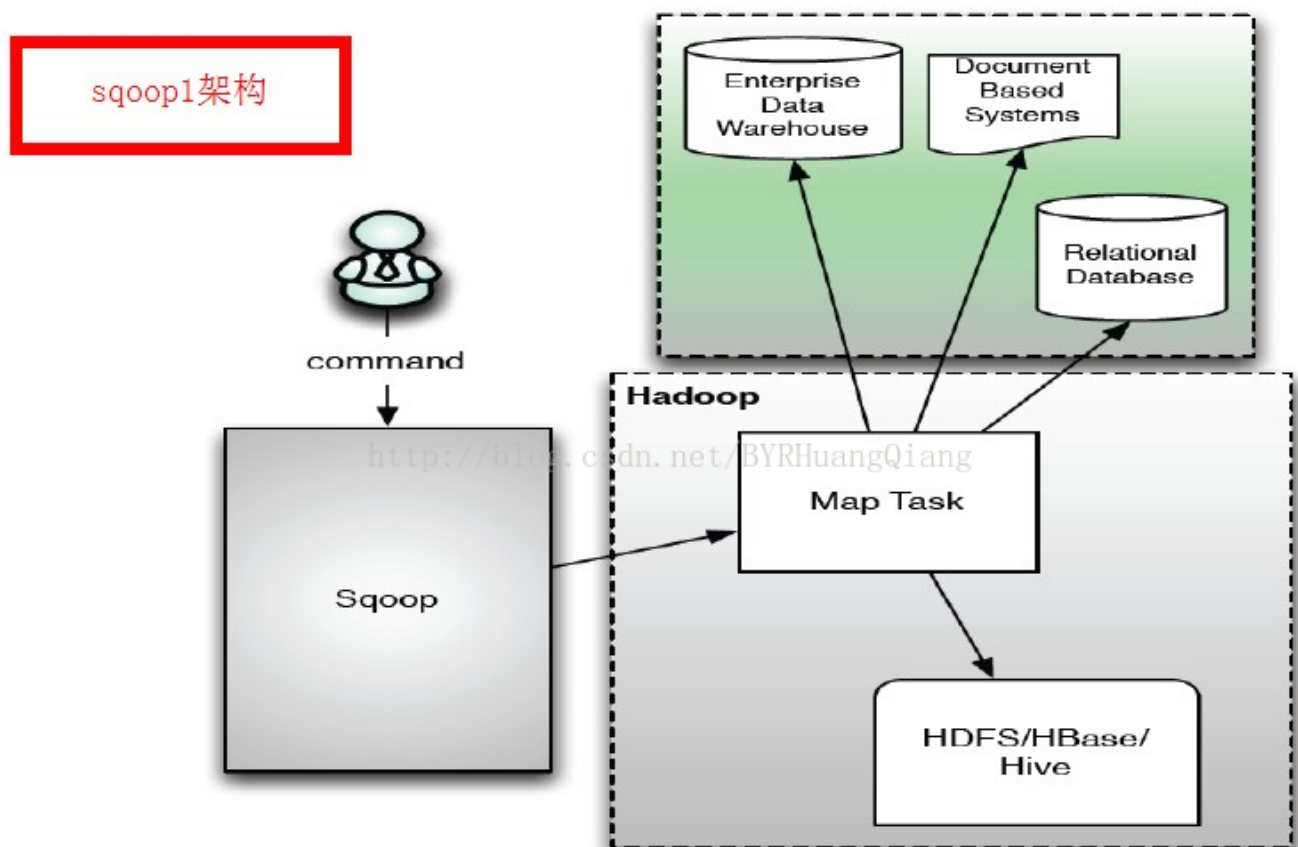
2.2 Sqoop1 和 Sqoop2 比较

2.2.1 Sqoop1 和 Sqoop2 异同

- 两个不同的版本，完全不兼容
- 版本号划区别，Apache 版本：1.4.x(Sqoop1); 1.99.x(Sqoop2) CDH 版本：
Sqoop-1.4.3-cdh4(Sqoop1); Sqoop2-1.99.2-cdh4.5.0 (Sqoop2)
- Sqoop2 比 Sqoop1 的改进
 - (1) 引入 Sqoop server，集中化管理 connector 等
 - (2) 多种访问方式：CLI, Web UI，REST API
 - (3) 引入基于角色的安全机制

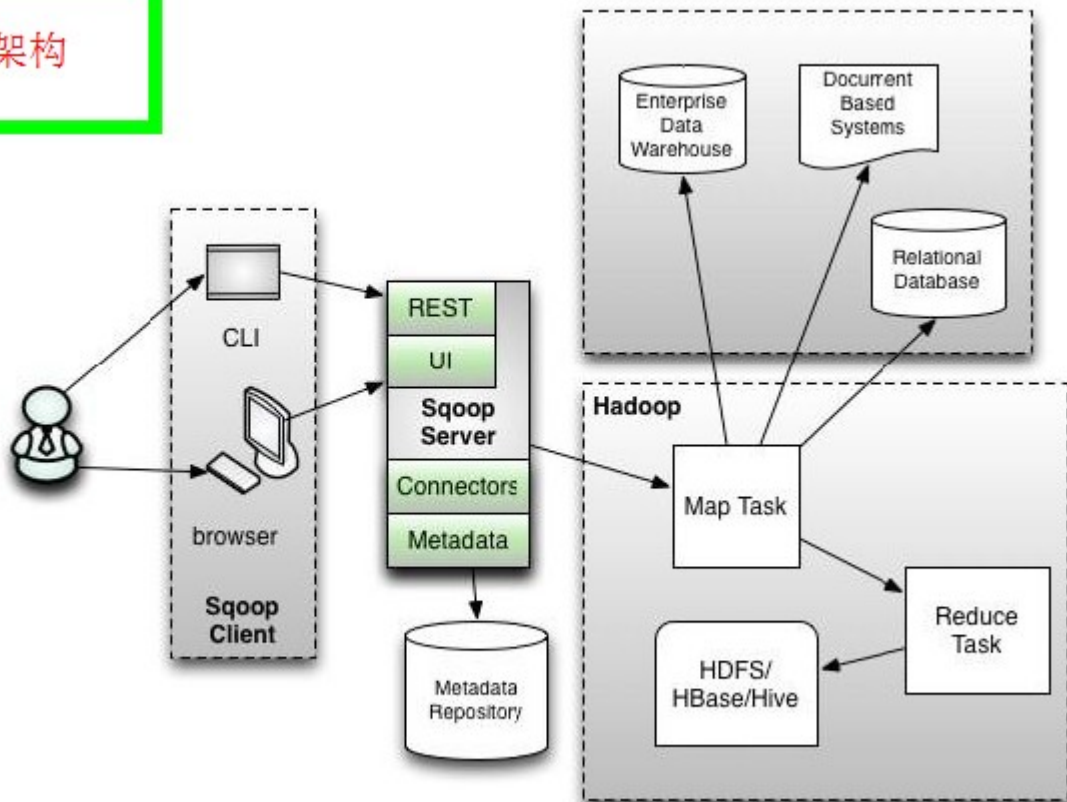
2.2.2 Sqoop1 与 Sqoop2 的架构图

Sqoop 架构图 1



Sqoop 架构图 2

sqoop2架构



2.2.3 Sqoop1 与 Sqoop2 的优缺点

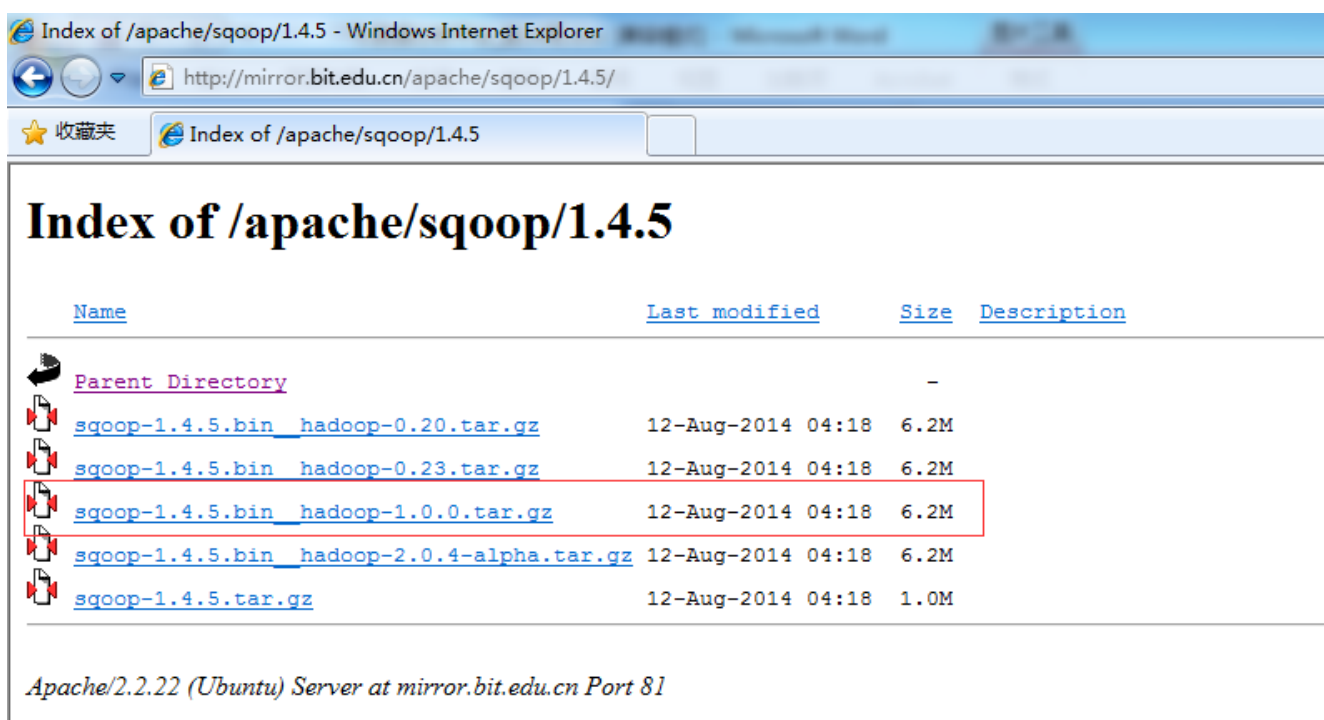
比较	Sqoop1	Sqoop2
架构	仅仅使用一个 Sqoop 客户端	引入了 Sqoop server 集中化管理 connector , 以及 rest api , web , UI , 并引入权限安全机制
部署	部署简单 , 安装需要 root 权限 , connector 必须符合 JDBC 模型	架构稍复杂 , 配置部署更繁琐
使用	命令行方式容易出错 , 格式紧耦合 , 无法支持所有数据类型 , 安全机制不够完善 , 例如密码暴漏	多种交互方式 , 命令行 , web UI , rest API , connctor 集中化管理 , 所有的链接安装在 Sqoop server 上 , 完善权限管理机制 , connector 规范化 , 仅仅负责数据的读写

比较	Sqoop1	Sqoop2
架构	仅仅使用一个 Sqoop 客户端	引入了 Sqoop server 集中化管理 connector，以及 rest api，web，UI，并引入权限安全机制
部署	部署简单，安装需要 root 权限，connector 必须符合 JDBC 模型	架构稍复杂，配置部署更繁琐
使用	命令行方式容易出错，格式紧耦合，无法支持所有数据类型，安全机制不够完善，例如密码暴漏	多种交互方式，命令行，web UI，rest API，connector 集中化管理，所有的链接安装在 Sqoop server 上，完善权限管理机制，connector 规范化，仅仅负责数据的读写

3 安装部署 Sqoop

3.1 下载 Sqoop

可以到 apache 基金 sqoop 官网 <http://hive.apache.org/>，选择镜像下载地址：<http://mirror.bit.edu.cn/apache/sqoop/> 下载一个稳定版本，如下图所示下载支持 Hadoop1.X 的 1.4.5 版本 gz 包：



也可以在/home/shiyanlou/install-pack 目录中找到该安装包，解压该安装包并把该安装包复制到/app 目录中

```
cd /home/shiyanlou/install-pack
tar -xzf sqoop-1.4.5.bin__hadoop-1.0.0.tar.gz
mv sqoop-1.4.5.bin__hadoop-1.0.0 /app/sqoop-1.4.5
ll /app
```

```
[shiyanlou@b393a04554e1 ~]$ cd /home/shiyanlou/install-pack
[shiyanlou@b393a04554e1 install-pack]$ tar -xzf sqoop-1.4.5.bin__hadoop-1.0.0.tar.gz
[shiyanlou@b393a04554e1 install-pack]$ mv sqoop-1.4.5.bin__hadoop-1.0.0 /app/sqoop-1.4.5
[shiyanlou@b393a04554e1 install-pack]$ ll /app
total 36
drwxrwxr-x 3 shiyanlou shiyanlou 4096 Jun  2 02:00 compile
drwxr-xr-x 19 shiyanlou shiyanlou 4096 Jun  6 14:22 hadoop-1.1.2
drwxrwxr-x 12 shiyanlou shiyanlou 4096 Jun  2 08:08 hadoop-2.2.0
drwxrwxr-x 8 shiyanlou shiyanlou 4096 Jun 10 03:17 hbase-0.96.2
drwxrwxr-x 8 shiyanlou shiyanlou 4096 Jun  9 14:58 hive-0.12.0
drwxrwxr-x 5 shiyanlou shiyanlou 4096 Jun  2 01:31 lib
drwxrwxr-x 15 shiyanlou shiyanlou 4096 Jun 10 01:50 mahout-0.6
drwxr-xr-x 15 shiyanlou shiyanlou 4096 Jun 29 2014 pig-0.13.0
drwxr-xr-x 9 shiyanlou shiyanlou 4096 Aug  1 2014 sqoop-1.4.5
```

3.2 设置/etc/profile 参数

编辑/etc/profile 文件，加入 sqoop 的 Home 路径和在 PATH 加入 bin 的路径：

```
export SQOOP_HOME=/app/sqoop-1.4.5
export PATH=$PATH:$SQOOP_HOME/bin
```

```
export HBASE_HOME=/app/hbase-0.96.2
export PATH=$PATH:$HBASE_HOME/bin
export SQOOP_HOME=/app/sqoop-1.4.5
export PATH=$PATH:$SQOOP_HOME/bin
```

编译配置文件/etc/profile，并确认生效

```
source /etc/profile
echo $PATH
```

3.3 设置 bin/configure-sqoop 配置文件

修改 bin/configure-sqoop 配置文件

```
cd /app/sqoop-1.4.5/bin
sudo vi configure-sqoop
```

注释掉 HBase 和 Zookeeper 等检查（除非使用 HBase 和 Zookeeper 等 HADOOP 上的组件）

```
## Moved to be a runtime check in sqoop.
#if [ ! -d "${HBASE_HOME}" ]; then
# echo "Warning: $HBASE_HOME does not exist! HBase imports will fail."
# echo 'Please set $HBASE_HOME to the root of your HBase installation.'
#fi

## Moved to be a runtime check in sqoop.
#if [ ! -d "${HCAT_HOME}" ]; then
# echo "Warning: $HCAT_HOME does not exist! HCatalog jobs will fail."
# echo 'Please set $HCAT_HOME to the root of your HCatalog installation.'
#fi

#if [ ! -d "${ACCUMULO_HOME}" ]; then
# echo "Warning: $ACCUMULO_HOME does not exist! Accumulo imports will fail."
# echo 'Please set $ACCUMULO_HOME to the root of your Accumulo installation.'
#fi
#if [ ! -d "${ZOOKEEPER_HOME}" ]; then
# echo "Warning: $ZOOKEEPER_HOME does not exist! Accumulo imports will fail."
# echo 'Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.'
#fi
```

3.4 设置 conf/sqoop-env.sh 配置文件

如果不存在 sqoop-env.sh 文件，复制 sqoop-env-template.sh 文件，然后修改 sqoop-env.sh 配置文件

```
cd /app/sqoop-1.4.5/conf
cp sqoop-env-template.sh sqoop-env.sh
sudo vi sqoop-env.sh
```

```
[shiyantou@b393a04554e1 ~]$ cd /app/sqoop-1.4.5/conf
[shiyantou@b393a04554e1 conf]$ cp sqoop-env-template.sh sqoop-env.sh
[shiyantou@b393a04554e1 conf]$ ls
oraoop-site-template.xml  sqoop-env-template.cmd  sqoop-site-template.xml
sqoop-env.sh             sqoop-env-template.sh  sqoop-site.xml
[shiyantou@b393a04554e1 conf]$ sudo vi sqoop-env.sh
```

设置 hadoop 运行程序所在路径和 hadoop-*-core.jar 路径 (Hadoop1.X 需要配置)

```
#Set path to where bin/hadoop is available
export HADOOP_COMMON_HOME=/app/hadoop-1.1.2
#Set path to where hadoop-*-core.jar is available
export HADOOP_MAPRED_HOME=/app/hadoop-1.1.2
```

```
#Set path to where bin/hadoop is available
export HADOOP_COMMON_HOME=/app/hadoop-1.1.2

#Set path to where hadoop-*-core.jar is available
export HADOOP_MAPRED_HOME=/app/hadoop-1.1.2

#set the path to where bin/hbase is available
#export HBASE_HOME=
```

编译配置文件 sqoop-env.sh 使之生效

3.5 验证安装完成

输入如下命令验证是否正确安装 sqoop，如果正确安装则出现 sqoop 提示

```
sqoop help
```



```
[shiyanolou@b393a04554e1 conf]$ sqoop help
Warning: $HADOOP_HOME is deprecated.

15/06/10 06:02:19 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5
usage: sqoop COMMAND [ARGS]

Available commands:
codegen          Generate code to interact with database records
create-hive-table Import a table definition into Hive
eval            Evaluate a SQL statement and display the results
export          Export an HDFS directory to a database table
help            List available commands
import          Import a table from a database to HDFS
import-all-tables Import tables from a database to HDFS
job             Work with saved jobs
list-databases  List available databases on a server
list-tables    List available tables in a database
merge          Merge results of incremental imports
metastore       Run a standalone Sqoop metastore
version         Display version information

See 'sqoop help COMMAND' for information on a specific command.
[shiyanolou@b393a04554e1 conf]$
```

4 文件导入/导出

4.1 MySql 数据导入到 HDFS 中

如果没有安装 MySql，请参照第 8 课 3.1 进行安装

4.1.1 下载 MySql 驱动

到 MySql 官网进入下载页面：<http://dev.mysql.com/downloads/connector/j/>，选择所需要的版本进行下载，这里下载的 zip 格式的文件，然后在本地解压：

Generally Available (GA) Releases

Connector/J 5.1.34

Select Platform:

Platform Independent

Looking for previous GA versions?

Platform Independent (Architecture Independent), Compressed TAR Archive	5.1.34	3.6M	Download
(mysql-connector-java-5.1.34.tar.gz)	MD5: ad4c875c719247f8e8c41cd7f0609e00 Signature		
Platform Independent (Architecture Independent), ZIP Archive	5.1.34	3.9M	Download
(mysql-connector-java-5.1.34.zip)	MD5: 51a49bfdeb963d1be3ac7bd41ac7908f Signature		

We suggest that you use the MD5 checksums and GnuPG signatures to verify the integrity of the packages you download.

也可以在/home/shiyanolou/install-pack 目录中找到该安装包，把 MySql 驱动包使用如下命令放到 Sqoop 的 lib 目录下

```
cd /home/shiyanolou/install-pack
```

```
cp mysql-connector-java-5.1.22-bin.jar /app/sqoop-1.4.5/lib
```

```
[shiyantou@b393a04554e1 ~]$ cd /home/shiyantou/install-pack
[shiyantou@b393a04554e1 install-pack]$ ls
apache-maven-3.0.5-bin.tar.gz      hbase-0.96.2-hadoop1-bin.tar.gz
chukwa-0.6.0.tar.gz               hive-0.12.0-bin.tar.gz
class4                             jdk-7u55-linux-x64.tar.gz
class5                             mahout-distribution-0.6.tar.gz
class6                             MySQL-client-5.6.21-1.el6.x86_64.rpm
class7                             mysql-connector-java-5.1.22-bin.jar
class9                             MySQL-devel-5.6.21-1.el6.x86_64.rpm
eclipse-jee-luna-SR1-linux-gtk-x86_64.tar.gz MySQL-server-5.6.21-1.el6.x86_64.rpm
flume-1.5.2-bin.tar.gz            pig-0.13.0.tar.gz
hadoop-1.1.2-bin.tar.gz           protobuf-2.5.0.tar.gz
hadoop-2.2.0.tar.gz              sqoop-1.4.5.bin__hadoop-1.0.0.tar.gz
hadoop-eclipse-plugin-1.1.2.jar
[shiyantou@b393a04554e1 install-pack]$ cp mysql-connector-java-5.1.22-bin.jar /app/sqoop-1.4.5/lib
```

4.1.2 启动 MySQL 服务

查看 MySQL 服务并查看状态，如果没有启动则启动服务

sudo service mysql status

sudo service mysql start

```
[shiyantou@b393a04554e1 ~]$ sudo service mysql status
ERROR! MySQL is not running
[shiyantou@b393a04554e1 ~]$ sudo service mysql start
Starting MySQL. SUCCESS!
[shiyantou@b393a04554e1 ~]$
```

4.1.3 查看 MySQL 中的数据表

进入 MySQL 数据库，选择有数据的一张表查看内容，比较导出结果是否正确，输入如下命令：

mysql -uhive -phive

mysql>show databases;

mysql>use hive;

mysql>show tables;

mysql>select TBL_ID, CREATE_TIME, DB_ID, OWNER, TBL_NAME,TBL_TYPE from TBLS;

```
[shiyantou@b393a04554e1 ~]$ mysql -uhive -phive
Warning: Using a password on the command line interface can be insecure.
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 1
Server version: 5.6.21 MySQL Community Server (GPL)

Copyright (c) 2000, 2014, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| hive      |
| mysql    |
| performance_schema |
| test     |
+-----+
5 rows in set (0.00 sec)
```



```
mysql> select TBL_ID, CREATE_TIME, DB_ID, OWNER, TBL_NAME, TBL_TYPE from TBLS;
+-----+-----+-----+-----+-----+-----+
| TBL_ID | CREATE_TIME | DB_ID | OWNER   | TBL_NAME | TBL_TYPE |
+-----+-----+-----+-----+-----+-----+
|      6 | 1433862566 |     1 | shianlou | test     | MANAGED_TABLE |
+-----+-----+-----+-----+-----+-----+
1 row in set (0.00 sec)
```

4.1.4 把 MySql 数据导入到 HDFS 中

使用如下命令列出 MySql 中所有数据库：

```
sqoop list-databases --connect jdbc:mysql://hadoop:3306/ --username hive
--password hive
```

```
[shianlou@b393a04554e1 ~]$ sqoop list-databases --connect jdbc:mysql://hadoop:3306/ --us
ername hive --password hive
Warning: $HADOOP_HOME is deprecated.

15/06/10 06:08:02 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5
15/06/10 06:08:02 WARN tool.BaseSqoopTool: Setting your password on the command-line is i
nsecure. Consider using -P instead.
15/06/10 06:08:02 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset
:
information_schema
hive
mysql
performance_schema
test
```

使用如下命令把 hive 数据库 TBLS 表数据导入到 HDFS 中：

```
sqoop import --connect jdbc:mysql://hadoop:3306/hive --username hive --password
hive --table TBLS -m 1
```

- --username 数据库用户名
- --password 连接数据库密码
- --table 表名
- -m 1 表示 map 数

```
[shianlou@b393a04554e1 ~]$ sqoop import --connect jdbc:mysql://hadoop:3306/hive --username hive --password
hive --table TBLS -m 1
Warning: $HADOOP_HOME is deprecated.

15/06/10 06:08:49 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5
15/06/10 06:08:49 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider
using -P instead.
15/06/10 06:08:49 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
15/06/10 06:08:49 INFO tool.CodeGenTool: Beginning code generation
15/06/10 06:08:50 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `TBLS` AS t LIMIT 1
15/06/10 06:08:50 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `TBLS` AS t LIMIT 1
15/06/10 06:08:50 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /app/hadoop-1.1.2
Note: /tmp/sqoop-shianlou/compile/3f94602049e59fff3c8c750bcc13f23f/TBLS.java uses or overrides a deprecate
d API.
Note: Recompile with -Xlint:deprecation for details.
15/06/10 06:08:52 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-shianlou/compile/3f94602049e59
fff3c8c750bcc13f23f/TBLS.jar
15/06/10 06:08:53 WARN manager.MySQLManager: It looks like you are importing from mysql.
15/06/10 06:08:53 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
15/06/10 06:08:53 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
15/06/10 06:08:53 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
15/06/10 06:08:53 INFO mapreduce.ImportJobBase: Beginning import of TBLS
15/06/10 06:08:54 INFO db.DBInputFormat: Using read committed transaction isolation
15/06/10 06:08:55 INFO mapred.JobClient: Running job: job_201506100247_0005
15/06/10 06:08:56 INFO mapred.JobClient: map 0% reduce 0%
15/06/10 06:09:07 INFO mapred.JobClient: map 100% reduce 0%
15/06/10 06:09:08 INFO mapred.JobClient: Job complete: job_201506100247_0005
15/06/10 06:09:08 INFO mapred.JobClient: Counters: 18
```

4.1.5 查看导出结果

使用如下命令查看导出到 HDFS 结果，文件路径在当前用户 hadoop 目录下增加了 TBLS 表目录，查看 part-m-00000 文件：

```
hadoop fs -ls /user/shiyanlou/TBLS
```

```
hadoop fs -cat /user/shiyanlou/TBLS/part-m-00000
```

```
[shiyanlou@b393a04554e1 ~]$ hadoop fs -ls /user/shiyanlou/TBLS
Found 3 items
-rw-r--r--  1 shiyanlou supergroup          0 2015-06-10 06:09 /user/shiyanlou/TBLS/_SUCCESS
drwxr-xr-x  - shiyanlou supergroup          0 2015-06-10 06:08 /user/shiyanlou/TBLS/_logs
-rw-r--r--  1 shiyanlou supergroup        60 2015-06-10 06:09 /user/shiyanlou/TBLS/part-m-00000
[shiyanlou@b393a04554e1 ~]$ hadoop fs -cat /user/shiyanlou/TBLS/part-m-00000
6,1433862566,1,0,shiyanlou,0,6,test,MANAGED_TABLE,null,null
[shiyanlou@b393a04554e1 ~]$
```

4.2 MySql 数据导入到 Hive 中

4.2.1 启动 metastore 和 hiveserver

在使用 hive 之前需要启动 metastore 和 hiveserver 服务，通过如下命令启用：

```
hive --service metastore &
```

```
hive --service hiveserver &
```

启动用通过 jps 命令可以看到两个进行运行在后台

```
[shiyanlou@b393a04554e1 ~]$ hive --service metastore &
[1] 20988
[shiyanlou@b393a04554e1 ~]$ Starting Hive Metastore Server
15/06/10 06:20:39 WARN conf.HiveConf: DEPRECATED: Configuration property hive.metastore.local no longer has any effect. Make sure to provide a valid value for hive.metastore.uris if you are connecting to a remote metastore.

[shiyanlou@b393a04554e1 ~]$ hive --service hiveserver &
[2] 21102
[shiyanlou@b393a04554e1 ~]$ Starting Hive Thrift Server
15/06/10 06:20:45 WARN conf.HiveConf: DEPRECATED: Configuration property hive.metastore.local no longer has any effect. Make sure to provide a valid value for hive.metastore.uris if you are connecting to a remote metastore.

[shiyanlou@b393a04554e1 ~]$ jps
11560 SecondaryNameNode
20988 RunJar
11801 TaskTracker
11313 NameNode
11433 DataNode
21102 RunJar
11674 JobTracker
21217 Jps
[shiyanlou@b393a04554e1 ~]$
```

4.2.2 从 MySql 导入表数据到 Hive 中

使用如下命令到把 MySql 中 TBLS 表数据导入到 Hive 中：

```
sqoop import --connect jdbc:mysql://hadoop:3306/hive --username hive --password
```

```
hive --table TBLS --hive-table MySql2Hive --hive-import -m 1
```

- --username 为 mysql 中的数据库连接用户名
- --password 为 mysql 中的数据库连接密码

- --table 为导出表
- --hive-table test1 为导出表在 Hive 中的名称
- -m 1 表示 map 数

```
[shiyanolou@b393a04554e1 ~]$ sqoop import --connect jdbc:mysql://hadoop:3306/hive --username hive --password
hive --table TBLS --hive-table MySql2Hive --hive-import -m 1
Warning: $HADOOP_HOME is deprecated.
15/06/10 06:40:22 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5
15/06/10 06:40:22 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider
using -P instead.
15/06/10 06:40:22 INFO tool.BaseSqoopTool: Using Hive-specific delimiters for output. You can override
15/06/10 06:40:22 INFO tool.BaseSqoopTool: delimiters with --fields-terminated-by, etc.
15/06/10 06:40:22 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
15/06/10 06:40:22 INFO tool.CodeGenTool: Beginning code generation
15/06/10 06:40:23 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `TBLS` AS t LIMIT 1
15/06/10 06:40:23 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `TBLS` AS t LIMIT 1
15/06/10 06:40:23 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /app/hadoop-1.1.2
Note: /tmp/sqoop-shiyanolou/compile/95af0b02cfc1808ce206097c17cba222/TBLS.java uses or overrides a deprecate
d API.
Note: Recompile with -Xlint:deprecation for details.
15/06/10 06:40:25 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-shiyanolou/compile/95af0b02cfc18
08ce206097c17cba222/TBLS.jar
15/06/10 06:40:25 WARN manager.MySQLManager: It looks like you are importing from mysql.
15/06/10 06:40:25 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
15/06/10 06:40:25 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
15/06/10 06:40:25 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
15/06/10 06:40:25 INFO mapreduce.ImportJobBase: Beginning import of TBLS
15/06/10 06:40:27 INFO db.DBInputFormat: Using read committed transaction isolation
15/06/10 06:40:27 INFO mapred.JobClient: Running job: job_201506100247_0011
15/06/10 06:40:28 INFO mapred.JobClient: map 0% reduce 0%
15/06/10 06:40:38 INFO mapred.JobClient: map 100% reduce 0%
15/06/10 06:40:40 INFO mapred.JobClient: Job complete: job_201506100247_0011
15/06/10 06:40:40 INFO mapred.JobClient: Counters: 18
```

```
15/06/10 06:30:47 INFO mapreduce.ImportJobBase: Transferred 60 bytes in 15.842 seconds (3.7874 bytes/sec)
15/06/10 06:30:47 INFO mapreduce.ImportJobBase: Retrieved 1 records.
15/06/10 06:30:47 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `TBLS` AS t LIMIT 1
15/06/10 06:30:47 INFO hive.HiveImport: Removing temporary files from import process: hdfs://hadoop:9000/us
er/shiyanolou/TBLS/_logs
15/06/10 06:30:47 INFO hive.HiveImport: Loading uploaded data into Hive
15/06/10 06:30:50 INFO hive.HiveImport: 15/06/10 06:30:50 WARN conf.HiveConf: DEPRECATED: Configuration pro
perty hive.metastore.local no longer has any effect. Make sure to provide a valid value for hive.metastore.
uris if you are connecting to a remote metastore.
15/06/10 06:30:50 INFO hive.HiveImport: Logging initialized using configuration in jar:file:/app/hive-0.12.
0/lib/hive-common-0.12.0.jar!/hive-log4j.properties
15/06/10 06:30:53 INFO hive.HiveImport: OK
15/06/10 06:30:53 INFO hive.HiveImport: Time taken: 2.977 seconds
15/06/10 06:30:54 INFO hive.HiveImport: Loading data to table default.test1
15/06/10 06:30:54 INFO hive.HiveImport: Table default.test1 stats: [num_partitions: 0, num_files: 2, num_ro
ws: 0, total_size: 60, raw_data_size: 0]
15/06/10 06:30:54 INFO hive.HiveImport: OK
15/06/10 06:30:54 INFO hive.HiveImport: Time taken: 0.874 seconds
15/06/10 06:30:54 INFO hive.HiveImport: Hive import complete.
15/06/10 06:30:54 INFO hive.HiveImport: Export directory is empty, removing it.
```

从运行的日志可以看到，这个过程有两个阶段：

1. 第一个阶段是从 MySQL 中把数据到 HDFS 文件中
2. 第二个阶段是从 HDFS 中把数据写入到 MySQL 中

4.2.3 查看导出结果

登录 hive，在 hive 创建表并查看该表，命令如下：

hive

hive>show tables;

hive>desc MySql2Hive;

```

[shiyantou@b393a04554e1 ~]$ hive
15/06/10 06:41:28 WARN conf.HiveConf: DEPRECATED: Configuration property hive.metastore.
local no longer has any effect. Make sure to provide a valid value for hive.metastore.ur
is if you are connecting to a remote metastore.

Logging initialized using configuration in jar:file:/app/hive-0.12.0/lib/hive-common-0.1
2.0.jar!/hive-log4j.properties
hive> show tables;
OK
mysql2hive
test
Time taken: 1.911 seconds, Fetched: 2 row(s)
hive> desc MySql2Hive;
OK
tbl_id          bigint          None
create_time     int             None
db_id           bigint         None
last_access_time int             None
owner           string          None
retention       int             None
sd_id           bigint         None
tbl_name        string          None
tbl_type        string          None
view_expanded_text string          None
view_original_text string          None
Time taken: 0.198 seconds, Fetched: 11 row(s)

```