

大数据技术-第九章：Sqoop数据迁移
Sqoop模板命令



CONTENTS

01. Sqoop模板命令



01

Sqoop模板命令



» Sqoop模板命令

(1) 创建MySQL数据库和数据表。

创建sample数据库，在sample中创建student表，在student表中插入了3条数据。

```
[hadoop@master ~]$ mysql -uroot -p # 登录MySQL数据库
```

```
Enter password:
```

```
mysql> create database sample; # 创建sample库
```

```
Query OK, 1 row affected (0.00 sec)
```

```
mysql> use sample; # 使用sample库
```

```
Database changed
```

```
mysql> create table student(number char(9) primary key, name varchar(10));
```

```
Query OK, 0 rows affected (0.01 sec) # 创建student表，该数据表有number学号和name姓名两个字段
```

» Sqoop模板命令

```
mysql> insert into student values('01','zhangsan');  
Query OK, 1 row affected (0.05 sec)
```

向student表插入几条数据

```
mysql> insert into student values('02','lisi');  
Query OK, 1 row affected (0.01 sec)
```

```
mysql> insert into student values('03','wangwu');  
Query OK, 1 row affected (0.00 sec)
```

```
mysql> select * from student;
```

查询student表的数据

```
+-----+-----+  
| number | name  |  
+-----+-----+  
| 01     | zhangsan |  
| 02     | lisi    |  
| 03     | wangwu  |  
+-----+-----+
```

```
3 rows in set (0.00 sec)
```

» Sqoop模板命令

(2) 在Hive中创建sample数据库和student数据表。

```
[hadoop@master ~]$ hive          # 启动hive命令行
Logging initialized using configuration in jar:file:/usr/local/src/hive/lib/hive-common-1.1.0.jar!/hive-log4j.properties
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/src/hive/lib/hive-jdbc-1.1.0-standalone.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
hive> create database sample;      # 创建sample库
OK
```


Scoop模板命令

hive> show databases; # 查询所有数据库

OK

default

sample

Time taken: 0.178 seconds, Fetched: 2 row(s)

hive>

hive> use sample; # 使用sample库

OK

hive> create table student(number STRING, name STRING)

row format delimited

fields terminated by "|"

stored as textfile; # 创建student表

OK

hive> exit; # 退出hive命令行

» Sqoop模板命令

(3) 从MySQL导出数据，导入Hive。

需要说明该命令的以下几个参数：

- 1) --connect: MySQL数据库连接URL。
- 2) --username和--password: MySQL数据库的用户名和密码。
- 3) --table: 导出的数据表名。
- 4) --fields-terminated-by: Hive中字段分隔符。
- 5) --delete-target-dir: 删除导出目的目录。
- 6) --num-mappers: Hadoop执行Sqoop导入导出启动的map任务数。
- 7) --hive-import --hive-database: 导出到Hive的数据库名。
- 8) --hive-table: 导出到Hive的表名。

```
[hadoop@master ~]$ sqoop import --connect jdbc:mysql://master:3306/sample --username root --password Password123$ --table student --fields-terminated-by '|' --delete-target-dir --num-mappers 1 --hive-import --hive-database sample --hive-table student
```


» Sqoop模板命令

删除MySQL数据，student表中number为主键，添加信息导致主键重复，报错，所以删除表数据。

```
[hadoop@master ~]$mysql -u root -p #密码Password123$
```

```
mysql>use sample;
```

```
mysql>delete from student;
```

```
mysql>exit;
```

» Sqoop模板命令

(4) 从Hive导出数据，导入到MySQL。

需要说明该命令的以下几个参数：

- 1) --connect: MySQL数据库连接URL。
- 2) --username和--password: MySQL数据库的用户名和密码。
- 3) --table: 导出的数据表名。
- 4) --fields-terminated-by: Hive中字段分隔符。
- 5) --export-dir: Hive数据表在HDFS中的存储路径。

```
[hadoop@master ~]$ sqoop export --connect  
"jdbc:mysql://master:3306/sample?useUnicode=true&characterEncoding=utf-8" --  
username root --password Password123$ --table student --input-fields-terminated  
-by '|' --export-dir /user/hive/warehouse/recruitdata.db/student/*
```

Turing AI 万维图灵 | 大数据系列课程

大数据

BIG
DATA

智 / 能 / 科 / 技 放 / 眼 / 未 / 来

