

Ejercicio Sqoop: importar una tabla en MySQL a Hive

Creación de la tabla en MYSQL

Probamos la conexión con mysql

```
mysql -u root -p  
Password: cloudera
```

Vemos las bbdd que contiene

```
Show databases;
```

```
mysql> Show databases;  
+-----+  
| Database |  
+-----+  
| information_schema |  
| cm |  
| firehose |  
| hue |  
| metastore |  
| mysql |  
| nav |  
| navms |  
| oozie |  
| retail_db |  
| rman |  
| sentry |  
+-----+
```

Creamos en MYSQL la tabla que queremos importar en hive

```
create database pruebaadb;
```

Creamos una tabla con datos que luego importaremos a hive mediante sqoop

```
use pruebaadb;
```

```
create table tabla_prueba (nombre varchar(30), edad int);
```

comprobamos que se ha creado

```
show tables;
```

```
mysql> show tables;  
+-----+  
| Tables_in_pruebaadb |  
+-----+  
| tabla_prueba |  
+-----+
```

Importamos algunas filas

```

INSERT INTO tabla_prueba VALUES
("Alberto",22);

INSERT INTO tabla_prueba VALUES ("Luis",
23);

INSERT INTO tabla_prueba VALUES
("Pablo", 24);

INSERT INTO tabla_prueba VALUES
("Carlos", 25);

```

```
INSERT INTO tabla_prueba VALUES ("Alberto",22);
```

```
INSERT INTO tabla_prueba VALUES ("Luis",23);
```

```
INSERT INTO tabla_prueba VALUES ("Pablo",24);
```

```
INSERT INTO tabla_prueba VALUES ("Carlos",25);
```

Comprobamos que los datos se han insertado en la tabla

```
Select * from tabla_prueba;
```

```
Describe tabla_prueba;
```

```

mysql> Select * from tabla_prueba;
+-----+-----+
| nombre | edad |
+-----+-----+
| Alberto | 22 |
| Luis | 23 |
| Pablo | 24 |
| Carlos | 25 |
+-----+-----+
4 rows in set (0.00 sec)

mysql>
mysql> Describe tabla_prueba;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| nombre | varchar(30) | YES | | NULL | |
| edad | int(11) | YES | | NULL | |
+-----+-----+-----+-----+-----+-----+

```

Creación de la tabla en HIVE

Creamos la tabla en hive donde se importarán los datos que acabamos de crear

Accedemos a hive

```
hive
```

Creamos una base de datos para esta prueba y accedemos a ella

```
create database prueba_sqoop_hive;
```

```
use prueba_sqoop_hive;
```

Comprobamos que está en el warehouse de hive

```
hadoop fs -ls /user/hive/warehouse
```

Creamos la estructura de la table que contendrá los datos importados desde mysql con sqoop

```
Drop table tabla_prueba_hive;  
CREATE TABLE tabla_prueba_hive  
(nombre string,  
edad int  
)  
ROW FORMAT DELIMITED  
STORED AS TEXTFILE;
```

```
DROP TABLE tabla_prueba_hive;
```

```
CREATE TABLE tabla_prueba_hive
```

```
( nombre string,
```

```
edad int )
```

```
ROW FORMAT DELIMITED STORED AS TEXTFILE;
```

Comprobamos que se ha creado con éxito

```
Show tables;
```

```
hive> Show tables;
```

```
OK
```

```
tabla_prueba_hive
```

Importamos la tabla con SQOOP

Dado que la “bbdd” Accumulo no está configurada, abrimos un Shell y ejecutamos los siguientes comandos para evitar warnings molestos.

```
sudo mkdir /var/lib/accumulo
```

```
ACCUMULO_HOME='/var/lib/accumulo'
```

```
export ACCUMULO_HOME
```

En un Shell escribimos lo siguiente para ver que sqoop está conectado con nuestro mysql:

```
sqoop list-databases --connect jdbc:mysql://localhost --
```

```
username root --password cloudera
```

Ahora listamos la tabla “table_prueba” de la bbdd “pruebadb” que hemos creado en MySQL

```
sqoop list-tables --connect jdbc:mysql://localhost/pruebadb --
```

```
username root --password cloudera
```

```
[cloudera@quickstart ~]$ sqoop list-tables --connect jdbc:mysql://localhost/pruebadb --username root --password cloudera
21/06/11 02:37:44 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
21/06/11 02:37:44 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
21/06/11 02:37:44 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
tabla_prueba
```

Usando los argumentos de importación hive mostrados en las slides del curso, importar la tabla creada en Mysql en la estructura creada en hive. Usar como conector (jdbc:mysql://localhost/bbddMysql) y un solo mapper.

```
sqoop import --connect jdbc:mysql://localhost/pruebadb --
table tabla_prueba --username root --password cloudera -m 1 --
hive-import --hive-overwrite --hive-table
prueba_sqoop_hive.tabla_prueba
```

```
Map-Reduce Framework
  Map input records=4
  Map output records=4
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=190
  CPU time spent (ms)=1000
  Physical memory (bytes) snapshot=118722560
  Virtual memory (bytes) snapshot=1510182912
  Total committed heap usage (bytes)=60751872
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=38
21/06/11 02:39:22 INFO mapreduce.ImportJobBase: Transferred 38 bytes in 22.6048 seconds (1.6811 bytes/sec)
21/06/11 02:39:22 INFO mapreduce.ImportJobBase: Retrieved 4 records.
21/06/11 02:39:22 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `tabla_prueba` AS t LIMIT 1
21/06/11 02:39:22 INFO hive.HiveImport: Loading uploaded data into Hive

Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-1.1.0-cdh5.13.0.jar!/hive-log4j.properties
OK
Time taken: 1.942 seconds
Loading data to table prueba_sqoop_hive.tabla_prueba
chgrp: changing ownership of 'hdfs://quickstart.cloudera:8020/user/hive/warehouse/prueba_sqoop_hive.db/tabla_prueba': User does not belong to supergroup
Table prueba_sqoop_hive.tabla_prueba stats: [numFiles=1, numRows=0, totalSize=38, rawDataSize=0]
OK
Time taken: 0.473 seconds_
```