

SQOOP

Sqoop es una herramienta Open Source creada originalmente por Cloudera

Es utilizada para transferir datos entre RDBMs y Hadoop

Ahora es un proyecto perteneciente al Apache Software Foundation

Ejercicio Sqoop:

importar una tabla en MySQL a Hive Creación de la tabla en MYSQL

1. Probamos la conexión con mysql

a. Mysql -u root -p

b. Password: cloudera

```
[cloudera@quickstart ~]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 26
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, 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>
```

2. Vemos las bbdd que contiene

a. Show databases;

```
mysql> SHOW databases;
+-----+
| Database |
+-----+
| information_schema |
| cm        |
| firehose  |
| hue       |
| metastore |
| mysql     |
| nav       |
| navms     |
| oozie     |
| retail_db |
| rman      |
| sentry    |
+-----+
12 rows in set (0.00 sec)

mysql>
```

3. Creamos en MYSQL la table que queremos importar en hive

a. create database pruebadb;

```
mysql> CREATE DATABASE pruebadb;
Query OK, 1 row affected (0.01 sec)

mysql>
```

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

a. use pruebadb; b. create table tabla_prueba (nombre varchar(30), edad int);

```
hive> Drop table tabla_prueba hive;CREATE TABLE tabla_prueba hive ( nombre string, edad int ) ROW FORMAT DELIMITED STORED AS TEXTFILE;
OK
Time taken: 0.018 seconds
OK
Time taken: 4.687 seconds
hive>
```

5. comprobamos que se ha creado

a. show tables;

```
mysql> SHOW tables;
+-----+
| Tables_in_pruebadb |
+-----+
| tabla_prueba       |
+-----+
1 row in set (0.00 sec)

mysql>
```

6. Importamos algunas filas a.

```
mysql> INSERT INTO tabla_prueba VALUES ("ALberto",22);
Query OK, 1 row affected (0.01 sec)

mysql> INSERT INTO tabla_prueba VALUES ("Luis",23);
Query OK, 1 row affected (0.01 sec)

mysql> INSERT INTO tabla_prueba VALUES ("pablo",24);
Query OK, 1 row affected (0.00 sec)

mysql> INSERT INTO tabla_prueba VALUES ("jose",25);
Query OK, 1 row affected (0.01 sec)

mysql>
```

7. Comprobamos que los datos se han insertado en la tabla

a. Select * from tabla_prueba;

b. Describe tabla_prueba;

```
mysql> SELECT * FROM tabla_prueba;
+-----+-----+
| nombre | edad |
+-----+-----+
| ALberto | 22 |
| Luis | 23 |
| pablo | 24 |
| jose | 25 |
+-----+-----+
4 rows in set (0.01 sec)

mysql> DESCRIBE tabla prueba;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near 'DESCRIBE tabla_prueba' at line 1
mysql> DESCRIBE tabla_prueba;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| nombre | varchar(30) | YES | | NULL | |
| edad | int(11) | YES | | NULL | |
+-----+-----+-----+-----+-----+-----+
2 rows in set (0.02 sec)

mysql>
```

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

1. Accedemos a hive

a. hive

```
drwxrwxr-x 4 cloudera cloudera 4096 Oct 24 2017 workspace
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p
roperties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> █
```

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

a. create database prueba_sqoop_hive;

b. use prueba_sqoop_hive;

```
hive> CREATE DATABASE prueba_sqoop_hive;
OK
Time taken: 0.069 seconds
hive> USE prueba_sqoop_hive
> ;
OK
Time taken: 0.013 seconds
hive> █
```

3. Comprobamos que está en el warehouse de hive

a. hadoop fs -ls /user/hive/warehouse

```
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse
Found 2 items
drwxrwxrwx - cloudera supergroup 0 2022-04-06 13:31 /user/hive/wareho
use/cursohivedb.db
drwxrwxrwx - cloudera supergroup 0 2022-04-07 14:41 /user/hive/wareho
use/prueba_sqoop_hive.db
[cloudera@quickstart ~]$ █
```

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

```
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 ("Santiago", 26); a.

5. Comprobamos que se ha creado con éxito

a. Show tables;ç

```
mysql> SHOW tables;
+-----+
| Tables_in_pruebadb |
+-----+
| tabla_prueba       |
+-----+
1 row in set (0.00 sec)

mysql>
```

Importamos la tabla con SQOOP

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

a. `sudo mkdir /var/lib/accumulo`

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

c. `export ACCUMULO_HOME`

```
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ sudo mkdir /var/lib/accumulo
[cloudera@quickstart ~]$ ACCUMULO_HOME='/var/lib/accumulo'
[cloudera@quickstart ~]$ export ACCUMULO_HOME
[cloudera@quickstart ~]$
```

2. En un Shell escribimos lo siguiente para ver que sqoop está conectado con nuestro mysql: a. `sqoop list-databases --connect jdbc:mysql://localhost --username root --password cloudera`

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

a. `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
22/04/08 11:26:35 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
22/04/08 11:26:35 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
22/04/08 11:26:36 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
tabla_prueba
[cloudera@quickstart ~]$
```

4. 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.

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