

Máster en Big Data

Tecnologías de Almacenamiento

10. Hands-On: Importar datos con Sqoop

Nombre: Àlex balló

Curso:2019

Índice

1. Introducción.....	4
2. Entorno	4
3. Explorar los datos origen.....	4
4. Uso de Sqoop	5

1. Introducción

El objetivo de este Hands-On es ver un ejemplo de uso de Sqoop para importar datos de una base de datos MySQL

2. Entorno

Para este Hands On, utilizaremos la máquina virtual desplegada en Hands-On anteriores llamada Desarrollo_Hadoop y todo será ejecutado vía shell

3. Explorar los datos origen

Lo primero que haremos es login en MySQL para ejecutar las instrucciones mediante Shell. Los datos que exploraremos son los relativos a base de datos movielens. Por todo ello nos conectaremos mediante el comando:

```
mysql --user=training --password=training movielens
```

- a) Revisa la estructura de la tabla movie y observa un ejemplo de datos de 5 registros

```
mysql> select * from movie limit 5;
```

id	name	year
1	Toy Story	1995
2	Jumanji	1995
3	Grumpier Old Men	1995
4	Waiting to Exhale	1995
5	Father of the Bride Part II	1995

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

- b) Revisa la estructura de la tabla movierating y observa un ejemplo de datos de 5 registros

```
mysql> select * from movierating limit 5;
```

userid	movieid	rating
1	1193	5
1	661	3
1	914	3
1	3408	4
1	2355	5

```
5 rows in set (0.02 sec)
```

4. Uso de Sqoop

a) Muestra la ayuda de Sqoop

```
[training@localhost ~]$ sqoop help
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.

```
[training@localhost ~]$ █
```

b) Lista las bases de datos de MySQL desde Sqoop

```
[training@localhost ~]$ sqoop list-databases --username training \
--password training \
--connect jdbc:mysql://localhost
19/05/13 14:04:31 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
19/05/13 14:04:31 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
information_schema
dualcore
hue
metastore
movielens
mysql
test
training
-
```

c) Lista las tablas de la base de datos movielens

```
[training@localhost ~]$ sqoop list-tables --username training \
--password training \
--connect jdbc:mysql://localhost/movielens
19/05/13 14:02:29 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
19/05/13 14:02:29 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
genre
movie
moviegenre
movierating
occupation
user
```

d) Importar la tabla movie y movierating

```
[training@localhost ~]$ sqoop import --username training \
--password training \
--connect jdbc:mysql://localhost/movielens \
--table movie \
--table movierating █
```

- e) Lista los archivos importados. Según la nomenclatura de los archivos resultantes, ¿ Que característica tiene el Job de Sqoop que se ha ejecutado?

```
Found 9 items
drwxr-xr-x - training supergroup      0 2019-03-25 15:26 averagelength
drwxr-xr-x - training supergroup      0 2019-03-25 14:52 averagelengthFinal
drwxr-xr-x - training supergroup      0 2019-04-08 13:43 combinerIpCount
drwxr-xr-x - training supergroup      0 2019-04-08 14:55 ipCountPartitioner
drwxr-xr-x - training supergroup      0 2019-03-20 16:57 ipcount
drwxr-xr-x - training supergroup      0 2019-05-13 14:12 movie
drwxr-xr-x - training supergroup      0 2019-05-13 14:19 movierating
drwxr-xr-x - training supergroup      0 2019-03-20 15:53 shakespeare
-rw-r--r-- 1 training _supergroup 504941532 2019-03-20 15:56 weblog
```