

## Práctica 2: Importación de datos con Sqoop

Para trabajar con la consola de la base de datos MySQL hay que situarse en node1: `ssh root@node1`  
`mysql -u root -p` contraseña: `hadoop123`

```
[root@node1 ~]# mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 95
Server version: 5.7.24 MySQL Community Server (GPL)

Copyright (c) 2000, 2018, 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.
```

### Consulta de bases de datos, tablas y campos

`mysql> show databases;`

```
mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| amon        |
| hive        |
| hue         |
| movielens   |
| mysql       |
| nav         |
| navms       |
| oozie       |
| performance_schema |
| rman        |
| scm         |
| sentry      |
| sys         |
+-----+
14 rows in set (0.01 sec)
```

```
mysql> use movielens;
```

```
mysql> use movielens;  
Reading table information for completion of table and column names  
You can turn off this feature to get a quicker startup with -A  
  
Database changed
```

```
mysql> show tables;
```

```
mysql> show tables;  
+-----+  
| Tables_in_movielens |  
+-----+  
| Artists              |  
| genre                |  
| movie                |  
| moviegenre           |  
| movierating          |  
| occupation           |  
| user                 |  
| users                |  
+-----+  
8 rows in set (0.00 sec)
```

```
mysql> select * from movie limit 10;
```

```
mysql> select * from movie limit 10;  
+-----+  
| 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 |  
| 6  | Heat                               | 1995 |  
| 7  | Sabrina                            | 1995 |  
| 8  | Tom and Huck                       | 1995 |  
| 9  | Sudden Death                       | 1995 |  
| 10 | GoldenEye                          | 1995 |  
+-----+  
10 rows in set (0.00 sec)
```

```
mysql> select count(*) from movie;
```

```
mysql> select count(*) from movie;  
+-----+  
| count(*) |  
+-----+  
|      3881 |  
+-----+  
1 row in set (0.00 sec)
```



mysql> describe movie;

```
mysql> describe movie;
+-----+-----+-----+-----+-----+-----+
| Field | Type          | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| id    | int(11)       | NO   | PRI | 0        |       |
| name  | char(75)      | YES  |     | NULL     |       |
| year  | smallint(6)   | YES  |     | NULL     |       |
+-----+-----+-----+-----+-----+-----+
3 rows in set (0.00 sec)
```

Salimos de la consola de mySQL con quit;

```
mysql> quit;
Bye
[root@node1 ~]#
```

Importar la tabla con Sqoop

Desde node1 ejecutamos la siguiente sentencia:

```
sqoop import --connect jdbc:mysql://node1/movielens --username root \ --password hadoop123 --  
target-dir /user/root/peliculas --table movie \ --fields-terminated-by ';' -m 1 --driver  
com.mysql.jdbc.Driver
```

--connect → indica la conexión con la base de datos con la que interactúa.

--username → indica el nombre de usuario de la base de datos.

--password → indica la contraseña de la base de datos.

--table → indica la tabla donde se guardarán los datos.

--fields-terminated-by → guarda los valores separados por “;”.

-m x → indica el número de nodos (x) para la tarea.

--driver → indica el driver a utilizar.

NOTA: El comando debe ser ejecutado sin la backslash.

```
[root@node1 ~]# sqoop import --connect jdbc:mysql://node1/movielens --username root --password hadoop123  
--target-dir /user/root/peliculas --table movie --fields-terminated-by ';' -m 1 --driver com.mysql.jdbc  
.Driver  
Warning: /opt/cloudera/parcels/CDH-6.1.1-1.cdh6.1.1.p0.875250/bin/../lib/sqoop/./accumulo does not exist!  
Accumulo imports will fail.  
Please set $ACCUMULO_HOME to the root of your Accumulo installation.  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.1.1-1.cdh6.1.1.p0.875250/jars/slf4j-log4j1  
2-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jar:file:/opt/cloudera/parcels/CDH-6.1.1-1.cdh6.1.1.p0.875250/jars/log4j-slf4j-  
impl-2.8.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.
```

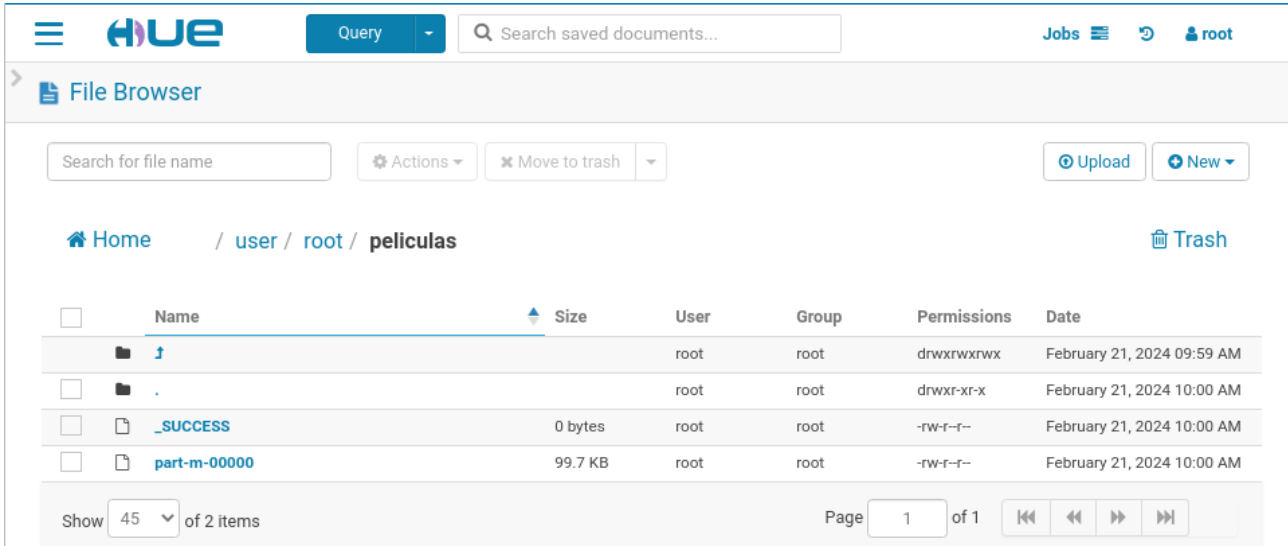
Aparece como resultado en la última línea:

INFO mapreduce.ImportJobBase: Retrieved 3881 records.

```
24/02/21 19:00:02 INFO mapreduce.ImportJobBase: Transferred 99.6602 KB in 30.2603 seconds (3.2934 KB/sec  
)  
24/02/21 19:00:02 INFO mapreduce.ImportJobBase: Retrieved 3881 records.
```

Nos indica el tiempo transcurrido y la tasa de transferencia, en la última línea el número de registros importados.

Podemos visualizar los datos desde Hue.



También lo podemos visualizar desde el cli con el comando “`hadoop fs -ls <ruta_hdfs>`”

```
[alumno@pasarela ~]$ hadoop fs -ls /user/root/peliculas
Found 2 items
-rw-r--r--  2 root root          0 2024-02-21 19:00 /user/root/peliculas/_SUCCESS
-rw-r--r--  2 root root    102052 2024-02-21 19:00 /user/root/peliculas/part-m-00000
```

Para visualizar el archivo podemos utilizar el comando “`hadoop fs -cat <ruta_fichero_hdfs>`”

```
[alumno@pasarela ~]$ FRANCISCO GARRIDO LARA ^C
[alumno@pasarela ~]$ hadoop fs -cat /user/root/peliculas/part-m-00000 | head -n 10
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
6;Heat;1995
7;Sabrina;1995
8;Tom and Huck;1995
9;Sudden Death;1995
10;GoldenEye;1995
```

Los archivos obtenidos son de texto delimitados por comas en este caso. Estando ya en HDFS pueden ser utilizados por el resto de aplicaciones de tratamiento de datos.

Recoge en un pantallazo el contenido de películas como muestra de que has hecho la práctica. Recuerda que se debe ver tu nombre en la imagen.

NOTA: Las tablas equivalen a los directorios y los datos equivalen a los ficheros.