# PRÁCTICA Nº 4.1

Cargar datos de MySQL a HDFS



- Módulo: Sistemas de Big Data.
- Nombre y apellidos: Alvaro Lucio-Villegas de Cea







## Índice

Ejercicio 1 Preparar la bbdd MySQL

3

Ejercicio 2 Cargar datos de MySQL a HDFS

7

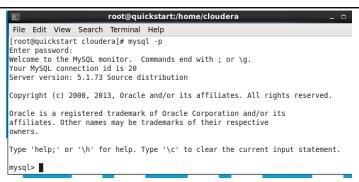




## Ejercicio 1 Preparar la bbdd MySQL

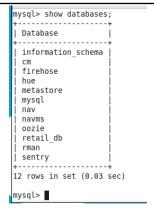
Acceder a MySql y si pide más permisos usar la segunda opción.

mysql mysql -u root -p



#### Mostramos la base de datos

mysql> show databases;



Crear una base de datos vacía.

mysql> create database mibd;

```
mysql> create database mibd;
Query OK, 1 row affected (0.01 sec)
mysql> ■
```





Nos ubicamos en la base de datos para trabajar en ella

mysql> use mibd;

mysql> use mibd; Database changed mysql> █

Creamos la tabla empleados en la bd

mysql> CREATE TABLE empleados(
id MEDIUMINT NOT NULL AUTO\_INCREMENT,
nombre CHAR(30) NOT NULL,
edad INTEGER(30),
salario INTEGER(30),
PRIMARY KEY (id) );

```
mysql> CREATE TABLE empleados(
-> id MEDIUMINT NOT NULL AUTO_INCREMENT,
-> nombre CHAR(30) NOT NULL,
-> edad INTEGER(30),
-> salario INTEGER(30),
-> PRIMARY KEY (id) );
Query OK, 0 rows affected (0.08 sec)

mysql> ■
```





#### Insertar datos en la bd

```
mysql> INSERT INTO empleados (nombre, edad, salario) VALUES ("Pedro Mora", 24, 21000), ("Maria Santos", 26, 24000), ("Juan Barba", 28, 25000), ("Benito Camela", 35, 28000), ("Monica Bono", 42, 30000), ("Rosa Melano", 43, 25000), ("Susana Oria", 45, 39000);
```

```
mysql> INSERT INTO empleados (nombre, edad, salario) VALUES
    -> ("Pedro Mora", 24, 21000),
    -> ("Maria Santos", 26, 24000),
    -> ("Juan Barba", 28, 25000),
    -> ("Benito Camela", 35, 28000),
    -> ("Monica Bono", 42, 30000),
    -> ("Rosa Melano", 43, 25000),
    -> ("Susana Oria", 45, 39000);
Query OK, 7 rows affected (0.01 sec)
Records: 7 Duplicates: 0 Warnings: 0
mysql> ■
```

#### Lee los datos con una SQL

#### mysql> select \* from empleados;





#### Se asignan todos los permisos al usuario

mysql> grant all privileges on \*.\* to 'root'@'localhost' IDENTIFIED BY 'cloudera' WITH GRANT OPTION;

```
mysql> grant all privileges on *.* to 'root'@'localhost' IDENTIFIED BY 'cloudera
' WITH
-> GRANT OPTION;
Query OK, 0 rows affected (0.05 sec)
mysql> ■
```

### Salir de MySql

mysql> exit;

mysql> exit Bye [root@quickstart cloudera]# **■** 





## Ejercicio 2 Cargar datos de MySQL a HDFS

Crear el directorio hdfs donde se volcara la tabla empleados

hdfs dfs -mkdir tablashdfsalv

[root@quickstart cloudera]# hdfs dfs -mkdir tablashdfsalv [root@quickstart cloudera]# ■

Comprueba que se ha creado con normalidad

hdfs dfs -ls

[root@quickstart cloudera]# hdfs dfs -ls
Found 1 items
drwxr-xr-x - root supergroup 0 2023-03-06 12:10 tablashdfsalv
[root@quickstart cloudera]# ■





Lanzar el Cliente Sqoop en línea de comando que ingestara la tabla correctamente.

sqoop import --connect jdbc:mysql://localhost/mibd --username root --password cloudera --table empleados --driver com.mysql.jdbc.Driver --target-dir /tablashdfsalv --fields-terminated-by ',' --lines-terminated-by '\n'

```
23/03/06 12:13:33 INFO mapreduce.Job: map 100% reduce 0%
23/03/06 12:13:34 INFO mapreduce.Job: Job job 1677098048375 0001 completed successfully
23/03/06 12:13:34 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=683644
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=377
                HDFS: Number of bytes written=162
                HDFS: Number of read operations=16
HDFS: Number of large read operations=0
                HDFS: Number of write operations=8
        Job Counters
                Launched map tasks=4
                Other local map tasks=4
                 Total time spent by all maps in occupied slots (ms)=19198
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=19198
                Total vcore-milliseconds taken by all map tasks=19198
                Total megabyte-milliseconds taken by all map tasks=19658752
        Map-Reduce Framework
                Map input records=7
                Map output records=7
                Input split bytes=377
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=619
                CPU time spent (ms)=4220
                Physical memory (bytes) snapshot=904130560
Virtual memory (bytes) snapshot=6314102784
                Total committed heap usage (bytes)=1092616192
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=162
23/03/06 12:13:34 INFO mapreduce.ImportJobBase: Transferred 162 bytes in 28.6588 seconds (5.6527 bytes/sec)
23/03/06 12:13:34 INFO mapreduce.ImportJobBase: Retrieved 7 records.
[root@quickstart cloudera]#
```

#### Que son cada parámetro del comando:

Claúsula	Significado		
import	Indica que el sentido de la información debe		
	ser desde la bbdd a Hadoop		
connect jdbc:mysql://localhost/mibd	Indica la cadena de conexión al tipo de bbdd,		
	el servidor (localhost) y la bbdd		
username rootpassword cloudera	Usuario y contraseña para conectarnos a la		
	bbdd. Nosotros estamos utilizando root, pero		
	si conocemos un usuario y contraseña con		
username root -P	acceso a esa bbdd y a las tablas que vamos a		
	utilizar, podríamos utilizar esas otras		
	credenciales.		
	ATTAICIÓN Is assessed abla se asses		
	ATENCIÓN lo recomendable es ponerusername root -P		
	porque de esa manera pide la clave por		
	teclado		
	teciado		
table empleados	Indica la tabla que vamos a importar a		
	Hadoop		
driver com.mysql.jdbc.Driver	Indica el driver a utilizar para conectarnos con		
, , ,	la bbdd		
target-dir/tablashdfs	Indica el directorio HDFS donde volcaremos la		
	información leída		
fields-terminated-by ','	Indica el carácter que se utilizará para separar		
	los valores, los campos		
lines-terminated-by '\n'	Indica el carácter utilizado para representar		
	los saltos de línea		





#### Comprueba que se han generado los ficheros en HDFS

hdfs dfs -ls /tablashdfsalv

hdfs dfs -cat /tablashdfsalv/part-m-0000[0-3]

```
[root@quickstart cloudera]# hdfs dfs -ls /tablashdfsalv
Found 5 items
-rw-r--r--
           1 root supergroup
                                      0 2023-03-06 12:13 /tablashdfsalv/ SUCCESS
-rw-r--r-- 1 root supergroup
                                     46 2023-03-06 12:13 /tablashdfsalv/part-m-00000
-rw-r--r-- 1 root supergroup
                                   47 2023-03-06 12:13 /tablashdfsalv/part-m-00001
-rw-r--r--
          1 root supergroup
                                    23 2023-03-06 12:13 /tablashdfsalv/part-m-00002
-rw-r--r--
          1 root supergroup
                                    46 2023-03-06 12:13 /tablashdfsalv/part-m-00003
[root@quickstart cloudera]#
```

Listamos los contenidos de los diferentes ficheros y podemos ver cómo se han segmentado los datos originales.

```
[root@quickstart cloudera]# hdfs dfs -cat /tablashdfsalv/part-m-00000
1,Pedro Mora,24,21000
2,Maria Santos,26,24000
[root@quickstart cloudera]# hdfs dfs -cat /tablashdfsalv/part-m-00001
3,Juan Barba,28,25000
4,Benito Camela,35,28000
[root@quickstart cloudera]# hdfs dfs -cat /tablashdfsalv/part-m-00002
5,Monica Bono,42,30000
[root@quickstart cloudera]# hdfs dfs -cat /tablashdfsalv/part-m-00003
6,Rosa Melano,43,25000
7,Susana Oria,45,39000
[root@quickstart cloudera]# hdfs dfs -cat /tablashdfsalv/part-m-00003
```