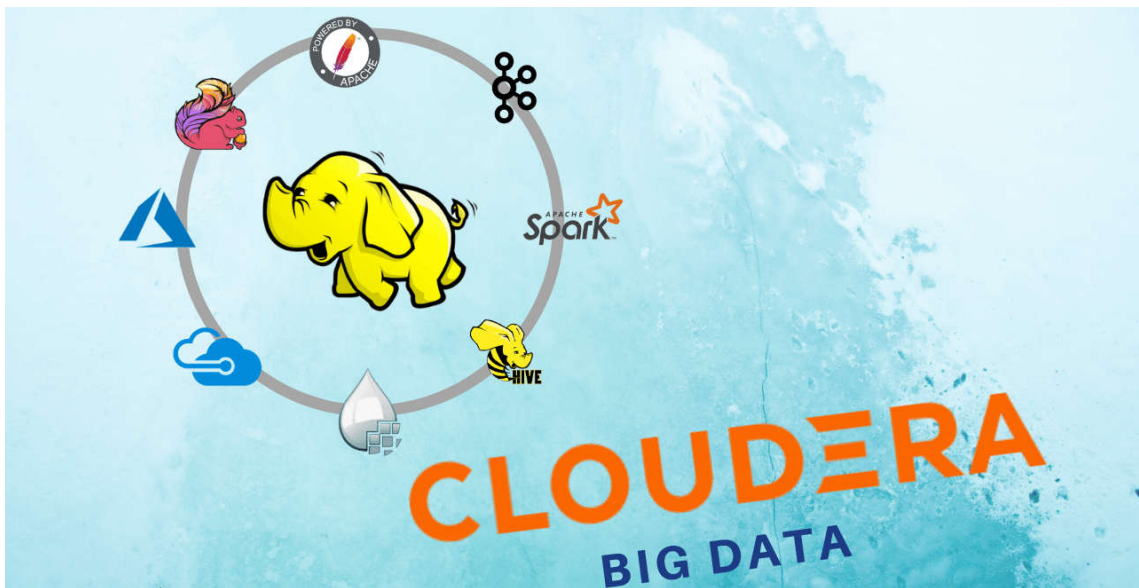


PRÁCTICA N° 4.2

Sqoop. Exportación de datos desde HDFS



- Alumno:Alvaro Lucio-Villegas de Cea



Índice

Ejercicio 1 – Preparar la bbdd MySQL	3
Ejercicio 2: Cargar datos de HDFS a MySQL	4
Ejemplo data-set de Telefonía	6



Ejercicio 1 – Preparar la bbdd MySQL

Usaremos la misma bases de datos del ejercicio anterior llamada “mibd” y sobre esta crearemos esta tabla completamente vacía. Esta será la tabla donde se cargaran los datos de la exportación.

```
CREATE TABLE copiaempleados(  
id MEDIUMINT NOT NULL AUTO_INCREMENT,  
nombre CHAR(30) NOT NULL,  
edad INTEGER(30),  
salario INTEGER(30),  
PRIMARY KEY (id) );
```

Ejecución y salida del comando

```
mysql> CREATE TABLE copiaempleados(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 (0.04 sec)
```

Ejercicio 2: Cargar datos de HDFS a MySQL

Una vez ya tengamos la tabla anterior creada correctamente realizaremos la exportación para ello nos tendremos que salir de sql y ejecutar el siguiente comando:

```
sqoop export --connect jdbc:mysql://localhost/mibd --username=root -P
--table=copiaempleados --export-dir=/tablashdfs -m=1
```

Ejecución del comando:

```
[cloudera@quickstart ~]$ sqoop export --connect jdbc:mysql://localhost/mibd --username=root -P --table=copiaempleados --export-dir=/tablashdfs -m=1
```

Salida del comando:

```
HDFS: Number of bytes written=0
HDFS: Number of read operations=13
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
Job Counters
  Launched map tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=2635
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=2635
  Total vcore-milliseconds taken by all map tasks=2635
  Total megabyte-milliseconds taken by all map tasks=2698240
Map-Reduce Framework
  Map input records=7
  Map output records=7
  Input split bytes=352
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=35
  CPU time spent (ms)=560
  Physical memory (bytes) snapshot=221581312
  Virtual memory (bytes) snapshot=1583423488
  Total committed heap usage (bytes)=273154048
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0
23/03/15 13:07:29 INFO mapreduce.ExportJobBase: Transferred 518 bytes in 15.9237 seconds (32.5301 bytes/sec)
23/03/15 13:07:29 INFO mapreduce.ExportJobBase: Exported 7 records.
[cloudera@quickstart ~]$
```

Explicación de los parámetros de los comandos.

Claúsula	Significado
export	Indica que el sentido de la información debe ser desde Hadoop a la bbdd
--connect jdbc:mysql://localhost/mibd	Indica la cadena de conexión al tipo de bbdd, el servidor (localhost) y la bbdd
--username root --password cloudera	En esta ocasión hemos utilizado la forma -P , pero podría haberse utilizado la otra versión, donde se suministra usuario y contraseña para conectarnos a la bbdd.
--username root -P	Igualmente, nosotros estamos utilizando root, pero si conocemos un usuario y contraseña con acceso a esa bbdd y a las tablas que vamos a utilizar, podríamos utilizar esas otras credenciales.
--table=copiaempleados	Tabla de destino donde se cargarán los datos leídos de los ficheros
--export-dir= / tablashdfs	Directorio HDFS que tiene los ficheros con los datos que deseamos exportar
-m=1	Número de tareas en paralelo que se utilizarán para exportar los datos

Comprobación de la importación.

```
[cloudera@quickstart ~]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 663
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> use mibd;
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> select * from copiaempleados;
+-----+-----+-----+-----+
| id | nombre          | edad | salario |
+-----+-----+-----+-----+
| 8  | Armando Bronca  | 24   | 21000   |
| 9  | Dolores Fuertes | 26   | 24000   |
| 10 | Javier Cidoncha | 28   | 25000   |
| 11 | Lorena          | 35   | 28000   |
| 12 | Miriam          | 42   | 30000   |
| 13 | Pedro           | 43   | 25000   |
| 14 | Juan            | 45   | 39000   |
+-----+-----+-----+-----+
7 rows in set (0.00 sec)

mysql> █
```

Ejemplo data-set de Telefonía

Carga de datos en hdfs

File Browser

Search for file name Actions

Home / user / cloudera / _sqoop

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	.		cloudera	cloudera	drwxr-xr-x	March 15, 2023 01:37 PM
<input type="checkbox"/>	WA_Fn-UseC_-Telco-Customer-Churn.csv	954.6 KB	cloudera	cloudera	drwxr-xr-x	March 15, 2023 01:41 PM

Show 45 of 1 items Page 1 of 1

Creación de tabla en la base de datos.

```
CREATE TABLE telefonica (
    customer_id VARCHAR(15) PRIMARY KEY,
    gender VARCHAR(10),
    senior_citizen VARCHAR(10),
    partner VARCHAR(30),
    dependents VARCHAR(30),
    tenure VARCHAR(10),
    phone_service VARCHAR(30),
    multiple_lines VARCHAR(30),
    internet_service VARCHAR(15),
    online_security VARCHAR(30),
    online_backup VARCHAR(30),
    device_protection VARCHAR(30),
    tech_support VARCHAR(30),
    streaming_tv VARCHAR(30),
    streaming_movies VARCHAR(30),
    contract VARCHAR(20),
    paperless_billing VARCHAR(30),
    payment_method VARCHAR(50),
    monthly_charges VARCHAR(30),
    total_charges VARCHAR(30),
    churn VARCHAR(3)
)
```

