

PREGUNTA 1.- Importar los datos de la tabla de MariaDB que contiene los datos de los departamentos (tabla departments) de la organización en la ruta /home/bigdata/pec1/departments en HDFS. Esta tabla al no poseer una columna de tipo entero hace imposible realizar la separación de los datos (split) y así aprovechar la concurrencia es por esta razón que es necesario que la tarea de importación se lleve a cabo mediante un único mapper.

Importation de MariaDB:

DB: employees

Table: departments

.. to HDFS Directory: /home/bigdata/pec1/departments

Num-mappers: 1

```
$ sqoop import --connect jdbc:mysql://bigdata:3306/employees --username root --password bigdata  
--table departments --target-dir /home/bigdata/pec1/departments --num-mappers 1
```

Resultado de la importación en HDFS:

```
$ hdfs dfs -cat /home/bigdata/pec1/departments/part-m-00000 | head
```

```
bigdata@bigdata:~$ hdfs dfs -cat /home/bigdata/pec1/departments/part-m-00000  
d009,Customer Service  
d005,Development  
d002,Finance  
d003,Human Resources  
d001,Marketing  
d004,Production  
d006,Quality Management  
d008,Research  
d007,Sales
```

PREGUNTA 2.- Importar los datos de la tabla de MariaDB que contiene los datos de los títulos (tabla titles) de la organización en la ruta /home/bigdata/pec1/titles en HDFS. Es necesario que al importar los datos los campos estén delimitados con el caracter ";" y a su vez los campos se encuentren encerrados entre comillas dobles P.e. "10001";"Senior Engineer";"1986-06-26";"9999-01-01"

Importacion de MariaDB:

DB: employees

Table: titles

.. to HDFS Directory: /home/bigdata/pec1/titles

Delimiter: ;

Fields enclosed by: "quotes"

```
$ sqoop import --connect jdbc:mysql://bigdata:3306/employees --username root --password bigdata
--table titles --target-dir /home/bigdata/pec1/titles --fields-terminated-by ';' --enclosed-by
','
```

```
Output directory hdfs://localhost:9000/home/bigdata/pec1/titles already exists
```

Resultado de la importación en HDFS: (imposible corregir '-->' sin borrar todo)

```
$ hdfs dfs -cat /home/bigdata/pec1/titles/part-m-00000 | head
```

```
bigdata@bigdata:~$ hdfs dfs -cat /home/bigdata/pec1/titles/part-m-00000 | head
'10001';'Senior Engineer';'1986-06-26';'9999-01-01'
'10002';'Staff';'1996-08-03';'9999-01-01'
'10003';'Senior Engineer';'1995-12-03';'9999-01-01'
'10004';'Engineer';'1986-12-01';'1995-12-01'
'10004';'Senior Engineer';'1995-12-01';'9999-01-01'
'10005';'Senior Staff';'1996-09-12';'9999-01-01'
'10005';'Staff';'1989-09-12';'1996-09-12'
'10006';'Senior Engineer';'1990-08-05';'9999-01-01'
'10007';'Senior Staff';'1996-02-11';'9999-01-01'
'10007';'Staff';'1989-02-10';'1996-02-11'
```

PREGUNTA 3.- Importar los datos de la tabla de MariaDB que contiene la información de los salarios de los empleados (tabla salaries) de la organización hacia nuestro Data Lake específicamente en Hive. Es necesario que al importar los datos los campos estén delimitados con el caracter "|" además de que únicamente sean importados los datos de los salarios mayores a 40000.

Importation de MariaDB:

DB: employees

Table: salaries

.. to Hive directory, /home/bigdata/salaries

Delimiter: "|"

where: salary > 40000

```
$ sqoop import --connect jdbc:mysql://bigdata:3306/employees --username root --password bigdata
--table salaries --hive-import --hive-overwrite --fields-terminated-by '|' --where
"salary>40000"
```

Resultado de la importación a HIVE: (también difícil corregir "|")

```
$ hdfs dfs -cat /user/bigdata/salaries/part-m-00000 | head
```

```
bigdata@bigdata:~$ hdfs dfs -cat /user/bigdata/salaries/part-m-00000 | head
10001000117001986-06-26001987-06-26
100010002102001987-06-26001988-06-25
100010006074001988-06-25001989-06-25
100010006596001989-06-25001990-06-25
100010006961001990-06-25001991-06-25
1000100071046001991-06-25001992-06-24
1000100074333001992-06-24001993-06-24
1000100075286001993-06-24001994-06-24
1000100075994001994-06-24001995-06-24
1000100076884001995-06-24001996-06-23
cat: Unable to write to output stream.
```

POR QUE NO ME APARECE AQUÍ table > "salaries" en Hive?? 😞

```
hive> SHOW tables;
OK
dept emp
Time taken: 0.079 seconds, Fetched: 1 row(s)
```

PREGUNTA 4.- Importar los datos de la tabla de MariaDB que contiene los datos de los empleados (tabla employees) de la organización en la ruta /home/bigdata/pec1/employees en HDFS. Es necesario que al importar los datos los campos estén delimitados con el caracter "|" y que como única condición que los empleados hayan sido contratados (hire_date) del año 1990 en adelante.

NOTA: Darle un vistazo a la función year() de SQL

Importacion de MariaDB:

DB: employees
Table: employees
.. to HDFS directory, /home/bigdata/pec1/employees
Delimiter: "|"
where: hire_date >= 1990;

(he tenido que crear el directorio HDFS 'employees1' porque no se borraba 'employees')

```
$ sqoop import --connect jdbc:mysql://bigdata:3306/employees --username root --password bigdata  
--table employees --target-dir /home/bigdata/pec1/employees1 --fields-terminated-by '|' --where  
"year(hire_date) > 1989"
```

Resultado de la importación en HDFS:

```
$ hdfs dfs -cat /home/bigdata/pec1/employees1/part-m-00000 | head
```

```
bigdata@bigdata:~$ hdfs dfs -cat /home/bigdata/pec1/employees1/part-m-00000 | head  
10008|1958-02-19|Saniya|Kalloufi|M|1994-09-15  
10011|1953-11-07|Mary|Sluis|F|1990-01-22  
10012|1960-10-04|Patricio|Bridgland|M|1992-12-18  
10016|1961-05-02|Kazuhiro|Cappelletti|M|1995-01-27  
10017|1958-07-06|Cristinel|Bouloucos|F|1993-08-03  
10019|1953-01-23|Lillian|Haddadi|M|1999-04-30  
10020|1952-12-24|Mayuko|Warwick|M|1991-01-26  
10022|1952-07-08|Shahaf|Famili|M|1995-08-22  
10024|1958-09-05|Suzette|Pettey|F|1997-05-19  
10026|1953-04-03|Yongqiao|Berztiss|M|1995-03-20  
cat: Unable to write to output stream.
```

Consulta en MariaDB:

```
> SELECT * FROM employees WHERE year(hire_date)>1989 ORDER BY hire_date LIMIT 100;
```

```
MariaDB [employees]> SELECT * FROM employees WHERE year(hire_date)>1989 ORDER BY hire_date LIMIT 100;  
+-----+-----+-----+-----+-----+-----+  
| emp_no | birth_date | first_name | last_name | gender | hire_date |  
+-----+-----+-----+-----+-----+-----+  
| 101299 | 1963-10-10 | Vivian | Horswill | F | 1990-01-01 |  
| 50654 | 1955-05-31 | Sreekrishna | Rodier | M | 1990-01-01 |  
| 99172 | 1956-03-25 | Chikako | Goldhammer | M | 1990-01-01 |  
| 254005 | 1953-02-27 | Souichi | Parfitt | F | 1990-01-01 |  
| 94548 | 1953-03-16 | Roddy | Dayana | F | 1990-01-01 |  
| 220057 | 1954-08-17 | Moty | Pauthner | M | 1990-01-01 |  
| 56998 | 1957-03-16 | Berni | Kandlur | F | 1990-01-01 |  
| 49447 | 1956-03-08 | Gianluca | Wallrath | M | 1990-01-01 |
```

PREGUNTA 5.- Exportar Los datos de los empleados (recientemente importados) ubicados en la ruta HDFS /home/bigdata/pec1/employees1 y exportarlos a la tabla en MariaDB employees-copy.
Si lo desean o porque les da problemas el nombre de la tabla "employees-copy" pueden crear una nueva en MariaDB llamada employees2 construida con la siguiente sentencia:

--> error con la version de MariaDB para consultar "employees-copy"

```
MariaDB [employees]> SELECT * FROM employees-copy limit 10;  
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MariaDB server version for the right syntax to use near '-copy limit 10' at line 1
```

---> creación de "employeescopy"

```
CREATE TABLE `employeescopy` (  
  `emp_no` int(11) NOT NULL,  
  `birth_date` date NOT NULL,  
  `first_name` varchar(14) NOT NULL,  
  `last_name` varchar(16) NOT NULL,  
  `gender` enum('M','F') NOT NULL,  
  `hire_date` date NOT NULL,  
  PRIMARY KEY (`emp_no`)  
) ENGINE=InnoDB DEFAULT CHARSET=latin1
```

Exportacion a MariaDB:

DB: employees
...from HDFS directory: /home/bigdata/pec1/employees1
.. to MariaDB table: employeescopy

```
sqoop export --connect "jdbc:mysql://bigdata:3306/employees" --username root --  
password bigdata --table employeescopy --export-dir /home/bigdata/pec1/employees1 --  
input-fields-terminated-by '/'
```

Resultado de la exportacion a MariaDB:

> SELECT * FROM employeescopy ORDER BY hire_date LIMIT 5;

```
MariaDB [employees]> SELECT * FROM employeescopy ORDER BY hire_date LIMIT 5;  
+-----+-----+-----+-----+-----+-----+  
| emp_no | birth_date | first_name | last_name | gender | hire_date |  
+-----+-----+-----+-----+-----+-----+  
| 21865 | 1964-10-24 | Lillian | Stiles | M | 1990-01-01 |  
| 24751 | 1953-01-24 | Aiman | Hainaut | M | 1990-01-01 |  
| 18199 | 1953-02-25 | Teruyuki | Sridhar | M | 1990-01-01 |  
| 18873 | 1961-03-05 | Sashi | Demeyer | F | 1990-01-01 |  
| 28869 | 1961-01-08 | Tomofumi | Rattan | M | 1990-01-01 |  
+-----+-----+-----+-----+-----+-----+  
5 rows in set (0.11 sec)
```