

Prácticas BigData

HIVE

1. Tablas internas

- Comprobar si hay bases de datos

```
show databases;
```

```
OK
```

```
+-----+
```

```
| database_name |
```

```
+-----+
```

```
| default      |
```

```
| ejemplo      |
```

```
| prueba       |
```

```
+-----+
```

```
3 rows selected (0,836 seconds)
```

- Nos conectamos a la Base de Datos de ejemplo

```
use ejemplo;
```

- Crear las siguientes tablas

```
CREATE TABLE IF NOT EXISTS empleados_internal
```

```
(
```

```
  name string,
```

```
  work_place ARRAY<string>,
```

```
  sex_age STRUCT<sex:string,age:int>,
```

```
  skills_score MAP<string,int>,
```

```
  depart_title MAP<STRING,ARRAY<STRING>>
```

```
)
```

```
COMMENT 'This is an internal table'
```

```
ROW FORMAT DELIMITED
```

```
FIELDS TERMINATED BY '|'
```

```
COLLECTION ITEMS TERMINATED BY ','
```

```
MAP KEYS TERMINATED BY ':'
```

- Lo cargamos con los datos del fichero empleados.txt que teneis en los recursos del curso.

```
LOAD DATA LOCAL INPATH '/home/curso/Downloads/empleados.txt'
OVERWRITE INTO TABLE empleados_internal;
```

```

Loading data to table ejemplo.empleados_internal
INFO : Loading data to table ejemplo.empleados_internal from
file:/home/cursos/Escritorio/employee.txt
Table ejemplo.empleados_internal stats: [numFiles=1, numRows=0,
totalSize=227, rawDataSize=0]
OK
INFO : Table ejemplo.empleados_internal stats: [numFiles=1, numRows=0,
totalSize=227, rawDataSize=0]
No rows affected (0,421 seconds)
0: jdbc:hive2://localhost:10000> select * from empleados_internal;
OK
+-----+-----+-----+-----+
+-----+-----+-----+-----+
| empleados_internal.name | empleados_internal.work_place |
empleados_internal.sex_age | empleados_internal.skills_score |
empleados_internal.depart_title |
+-----+-----+-----+-----+
+-----+-----+-----+-----+
| Michael | ["Montreal","Toronto"] | {"sex":"Male","age":30} |
{"DB":80} | {"Product":["Developer","Lead"]} |
| Will | ["Montreal"] | {"sex":"Male","age":35} | {"Perl":85}
| {"Product":["Lead"],"Test":["Lead"]} |
| Shelley | ["New York"] | {"sex":"Female","age":27} |
{"Python":80} | {"Test":["Lead"],"COE":["Architect"]} |
| Lucy | ["Vancouver"] | {"sex":"Female","age":57} |
{"Sales":89,"HR":94} | {"Sales":["Lead"]} |
+-----+-----+-----+-----+
+-----+-----+-----+-----+
4 rows selected (0,215 seconds)

```

- Comprobar que existe en el directorio warehouse de HIVE, dentro de la base de datos ejemplo. También lo podemos ver con HDFS

```

hdfs dfs -ls /user/hive/warehouse/ejemplo.db
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further
details.
Found 2 items
drwxrwxr-x - root supergroup 0 2015-06-11 11:15
/user/hive/warehouse/ejemplo.db/empleados_internal
drwxrwxr-x - root supergroup 0 2015-06-11 10:54 /user/hive/wareh

```

2. Tablas externas

- Creamos ahora una tabla externa. Hemos de asegurarnos de que tenemos el directorio /ejemplo, ya que es donde se van a quedar los datos.

```
CREATE EXTERNAL TABLE IF NOT EXISTS empleados_external
(
  name string,
  work_place ARRAY<string>,
  sex_age STRUCT<sex:string,age:int>,
  skills_score MAP<string,int>,
  depart_title MAP<STRING,ARRAY<STRING>>
)
COMMENT 'This is an external table'
ROW FORMAT DELIMITED
FIELDS TERMINATED BY '|'
COLLECTION ITEMS TERMINATED BY ','
MAP KEYS TERMINATED BY ':'
LOCATION '/ejemplo/empleados';
```

- Lo cargamos con los mismos datos

```
0: jdbc:hive2://localhost:10000> LOAD DATA LOCAL INPATH
'/home/curso/Desktop/empleados.txt' OVERWRITE INTO TABLE
empleados_external;

Loading data to table ejemplo.empleados_external

Table ejemplo.empleados_external stats: [numFiles=0, numRows=0,
totalSize=0, rawDataSize=0]

OK

INFO : Loading data to table ejemplo.empleados_external from
file:/home/curso/Escritorio/employee.txt

INFO : Table ejemplo.empleados_external stats: [numFiles=0, numRows=0,
totalSize=0, rawDataSize=0]

No rows affected (0,7 seconds)
```

- Probamos que estén las filas

```
0: jdbc:hive2://localhost:10000> select * from empleados_external;

OK

+-----+-----+-----+-----+
| empleados_external.name | empleados_external.work_place | empleados_external.sex_age | empleados_external.skills_score |
| empleados_external.depart_title |
```

+-----+-----+-----+-----+			
-----+-----+-----+-----+			
Michael	["Montreal","Toronto"]	{"sex":"Male","age":30}	
{"DB":80}	{"Product":["Developer","Lead"]}		
Will	["Montreal"]	{"sex":"Male","age":35}	{"Perl":85}
{"Product":["Lead"],"Test":["Lead"]}			
Shelley	["New York"]	{"sex":"Female","age":27}	
{"Python":80}	{"Test":["Lead"],"COE":["Architect"]}		
Lucy	["Vancouver"]	{"sex":"Female","age":57}	
{"Sales":89,"HR":94}	{"Sales":["Lead"]}		
+-----+-----+-----+-----+			
-----+-----+-----+-----+			
4 rows selected (0,137 seconds)			

- Comprobar que existen el directorio datos
- Hacer alguna SELECT por ejemplo para buscar al empleado "Lucy"
- Borrar la dos tablas
- Comprobar que ha borrado la interna, pero los datos de la externa permanecen.