

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

OK

Time taken: 1.945 seconds

- Hacemos una SELECT de los datos

```
select * from empleados_internal;
```

OK

```
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"]}
```

Time taken: 3.412 seconds, Fetched: 4 row(s))

- 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

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

Loading data to table ejemplo.empleados_external
OK
Time taken: 0.939 seconds
```

- Probamos que estén las filas

```
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}		
Will	["Montreal"]	{"sex":"Male","age":35}	85	
Shelley	["New York"]	{"sex":"Female","age":27}		
Lucy	["Vancouver"]	{"sex":"Female","age":57}		

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.