

# HIVE - EJERCICIO 1

## Creación y borrado de tablas

Para empezar, vamos a cargar en HDFS los datos del archivo "clients.csv".

Para ello, deberás copiar el archivo en HDFS como hiciste en la parte de HADOOP+HDFS.

A continuación, tras habernos conectado con el cliente `hive` o mediante `beeline`, creamos una base de datos llamada `iabd`:

```
create database if not exists iabd;
```

Nos conectamos a la base de datos que acabamos de crear:

```
use iabd;
```

A continuación, vamos a crear una tabla que almacene los valores del archivo csv:

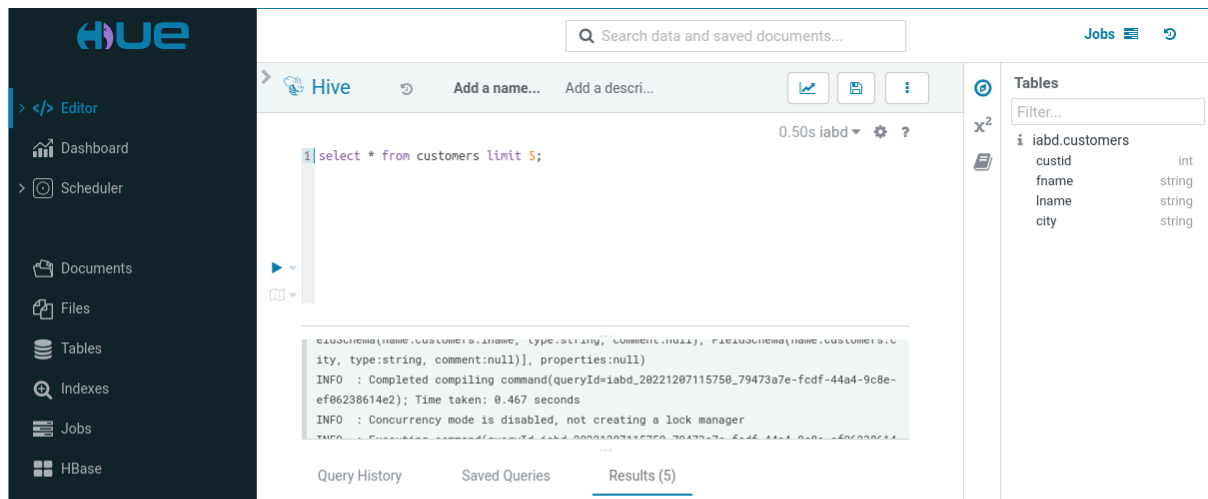
```
CREATE TABLE customers
(
  Name STRING COMMENT "Nombre",
  Situation STRING COMMENT "Tipo de cliente",
  country STRING
)
ROW FORMAT DELIMITED #Indica el formato de cada fila como delimitado (con un salto de línea)
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/iabd/hive/customer' ; #Ruta de HDFS donde se encuentran los datos
```

Y ya podemos realizar algunas consultas:

```
select * from customers limit 5;
```

```
select count(*) from customers;
```

Ahora prueba a hacer lo mismo utilizando HUE desde la máquina virtual de Cloudera:



En ocasiones necesitamos almacenar la salida de una consulta *Hive* en una nueva tabla. Las definiciones de las columnas de la nueva tabla se derivan de las columnas recuperadas en la consulta. Para ello, usaremos el comando *create table-as select*:

```
CREATE TABLE customers_new as SELECT * from customers;
```

En el caso de que la consulta falle por algún motivo, la tabla no se crearía.

Otra posibilidad es crear una tabla con la misma estructura que otra ya existente (pero sin datos):

```
CREATE TABLE customers2 LIKE customers;
```

En cualquier momento podemos obtener información de la tabla:

```
describe customers_new;
```

```
describe formatted customers_new;
```

Igual que las creamos, las podemos eliminar:

```
drop table customers_new;  
drop table customers2;
```

Si ejecutamos el comando `!tables` (o `show tables` en el cliente `hive`) veremos que ya no aparecen dichas tablas.

En el caso de que queramos eliminar una base de datos, de la misma manera que en SQL, ejecutaremos el comando `drop database iabd`. Si nuestra base de datos contiene tablas y las queremos borrar igualmente, entonces necesitamos un borrado en cascada mediante `drop database iabd cascade`.