IMPALA - EJERCICIO 1

Importación de datos desde fichero

Para empezar, descargamos el fichero data_berka.zip, lo descomprimimos y lo copiamos en la carpeta Workspace.

El archivo puede descargarse desde la siguiente ruta:

https://github.com/profemronda/archivosPDF/blob/main/data_berka.zip

Abrimos la terminal

```
cd workspace
cd data
ls
```

```
hdfs dfs -put account.asc .
```

Comprobamos que se ha copiado en nuestro sistema de ficheros:

```
hdfs dfs -ls
```

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

```
use iabd;
```

Ahora hay que darle permisos de escritura al directorio de nuestro hdfs, porque es donde va a almacenar los datos:

```
hdfs dfs -chmod 777 /user/cloudera
```

Ahora nos vamos a nuestra máquina Cloudera y abrimos HUE. Ahí seleccionamos IMPALA y creamos una nueva base de datos llamada DATA_BERKA.

Ahí crearemos una tabla importando los datos desde el fichero. La ruta del path es /user/cloudera/account.asc

Añadimos el separador del fichero, que en este caso es ';'

Veremos una previsualización. Tras pulsar en next, comprobaremos los tipos de los campos, y si todo es correcto, le damos a Submit.

Ya tendremos los datos importados en nuestro Impala.

Consulta de datos

Antes de comenzar, debemos repetir el paso anterior con el fichero disp.asc y client.asc

```
cd cloudera/data

hdfs dfs -put client.asc

hdfs dfs -put disp.asc
```

Desde Hue→Impala creamos una nueva tabla. En este caso, debemos indicarle el separador de campos (;) y también que el tipo de datos de birth_number es un string, ya que, si vemos el contenido del fichero, los datos de esa columna están entre comillas.

También debemos marcar la casilla “Store in default location”.

Una vez tengamos las tablas creadas, nos vamos a las consultas de Impala. Desde ahí probaremos diferentes consultas. En primer lugar, haremos un JOIN de las 3 tablas:

```
SELECT *
FROM account a, disp d, client c
WHERE a.account_id = d.account_id AND c.client_id = d.client_id
```

Podemos comprobar el resultado, así como las diferentes opciones que nos ofrece la interfaz gráfica. Entre otras opciones, podemos ver gráficas y exportar el resultado a un archivo csv.