PRÁCTICA Nº 2.5 Impala



Nombre y apellidos: Alvaro Lucio-Villegas de Cea





Índice

	_
Ejercicio 1	3
Ejercicio 2	4
Ejercicio 3	5

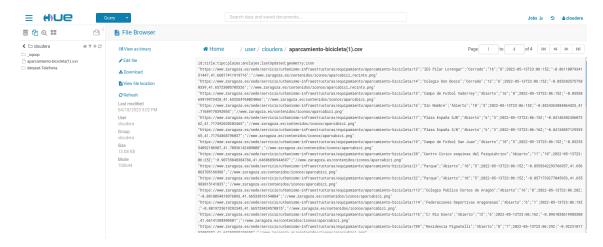




Ejercicio 1

Utilizando Impala mediante HUE, carga una tabla con los datos aparcamiento bicicleta, cuyo almacenamiento sea en formato CSV y el directorio donde se almacene se encuentre dentro de /user/cloudera para que pueda ser encontrado fácilmente.

Observamos la tabla una vez subida a HDFS.

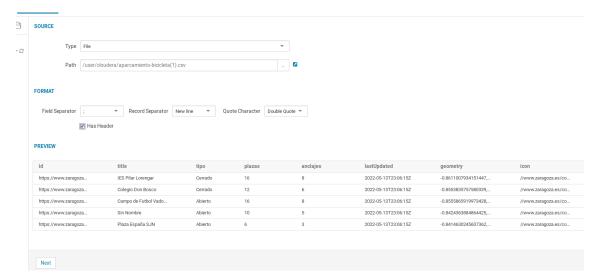


Asignamos permisos para que pueda escribir en el directorio correspondiente.

[cloudera@quickstart ~]\$ hdfs dfs -chmod 777 /user/cloudera [cloudera@quickstart ~]\$ ■

Comenzamos la creación de la tabla en Impala

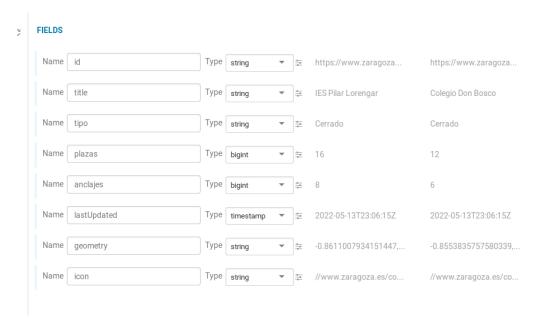
Asignamos el fichero donde se encuentran todos los registros. Tendremos que ajustar los parámetros para tener el formato de la tabla correctamente.



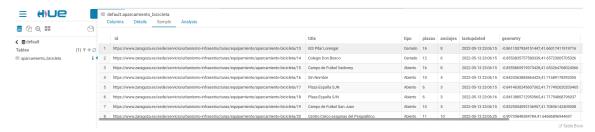




Una vez les asignamos los tipos correspondientes a cada columna.



Comprobamos que se ha creado correctamente la tabla.



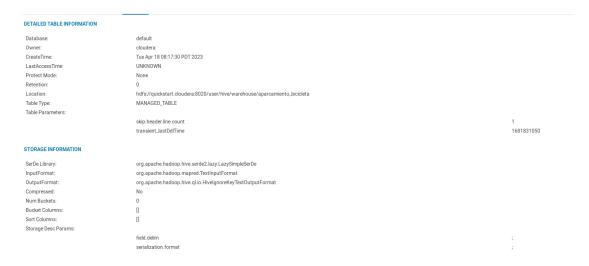




Ejercicio 2

Revisa mediante código los metadatos generados para ver cómo se tratan los datos almacenados en la tabla Impala.

Nos dirigimos a los detalles de la tabla desde la HUE.



Ejercicio 3

Haz una consulta que devuelva la información referente a un registro concreto.

Modifica el archivo CSV manualmente desde el sistema de archivos HDFS y repite la consulta para verificar que Impala trabaja directamente sobre el archivo de texto

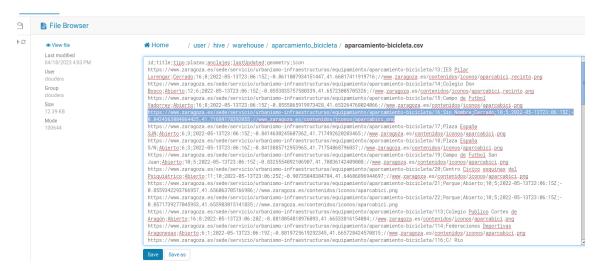
SELECT * FROM aparcamiento_bicicleta WHERE title like 'Sin Nombre'







Modificamos el fichero manualmente.



Ejecutamos de nuevo la consulta y vemos que ha cambiado.

