1. Crear las siguientes tablas externas en la base de datos f1 en hive:

a. driver\_results (driver\_forename, driver\_surname, driver\_nationality, points)

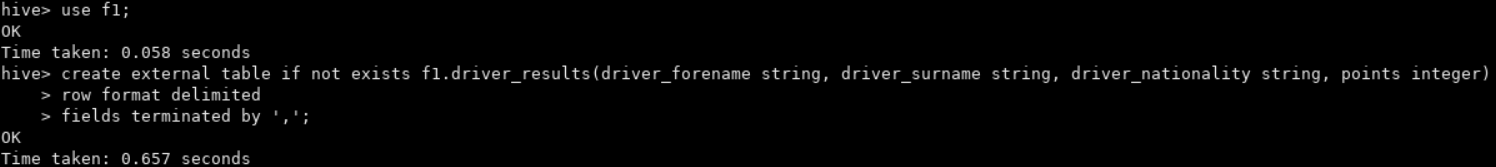
b. constructor\_results (constructorRef, cons\_name, cons\_nationality, url, points)

Creo la base de datos:

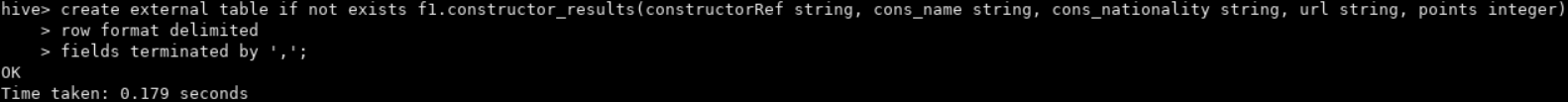
A screen shot of a computer

Description automatically generated

Creo la tabla driver\_results:



Creo la table constructor\_results:



1. En Hive, mostrar el esquema de driver\_results y constructor\_results

A computer screen with white text

Description automatically generated

A screen shot of a computer code

Description automatically generated

1. Crear un archivo .bash que permita descargar los archivos mencionados abajo e ingestarlos en HDFS:

**results.csv**

<https://data-engineer-edvai.s3.amazonaws.com/f1/results.csv>

**drivers.csv**

<https://data-engineer-edvai.s3.amazonaws.com/f1/drivers.csv>

**constructors.csv**

<https://data-engineer-edvai.s3.amazonaws.com/f1/constructors.csv>

Creamos el archivo ingest\_f1.sh

A computer screen with text and images

Description automatically generated

Le doy permisos de ejecución al bash

A black background with white text

Description automatically generated

1. Generar un archivo .py que permita, mediante Spark:
2. insertar en la tabla driver\_results los corredores con mayor cantidad de puntos en la historia.
3. insertar en la tabla constructor\_results quienes obtuvieron más puntos en el Spanish Grand Prix en el año 1991

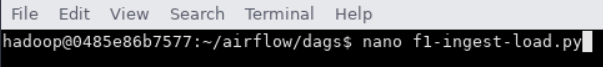
**Para resolver este ejercicio considero los 10 conductores/constructores con mayor cantidad de puntos**

**A screenshot of a computer

Description automatically generated**

1. Realizar un proceso automático en Airflow que orqueste los archivos creados en los puntos 3 y 4. Correrlo y mostrar una captura de pantalla (del DAG y del resultado en la base de datos).

Primero debo crear el DAG, por lo que genero ingest-load.py en la carpeta dags de airflow.



A screenshot of a computer

Description automatically generated

Una vez guardado busco el nuevo DAG en Airflow (f1-ingest-load).

A screenshot of a computer

Description automatically generated

Corremos el DAG y vemos que funcione bien.

A screenshot of a computer

Description automatically generated

Vemos el resultado en la base de datos.

A screenshot of a computer screen

Description automatically generated

A screen shot of a computer screen

Description automatically generated