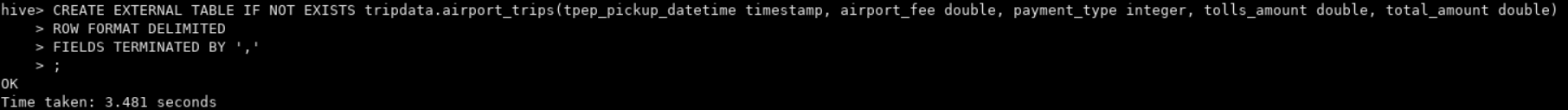
1. En Hive, crear la siguiente tabla (externa) en la base de datos tripdata:

airport\_trips(tpep\_pickup\_datetetime, airport\_fee, payment\_type, tolls\_amount, total\_amount)



1. En Hive, mostrar el esquema de airport\_trips

A screen shot of a computer

Description automatically generated

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

Yellow\_tripdata\_2021-01.parquet

(https://data-engineer-edvai.s3.amazonaws.com/yellow\_tripdata\_2021-01.parquet)

Yellow\_tripdata\_2021-02.parquet

(https://data-engineer-edvai.s3.amazonaws.com/yellow\_tripdata\_2021-02.parquet)

A screenshot of a computer

Description automatically generated

1. Crear un archivo .py que permita, mediante Spark, crear un data frame uniendo los viajes del mes 01 y mes 02 del año 2021 y luego Insertar en la tabla airport\_trips los viajes que tuvieron como inicio o destino aeropuertos, que hayan pagado con dinero.

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 screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Una vez guardado busco el nuevo DAG en Airflow (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

A screen shot of a black screen

Description automatically generated