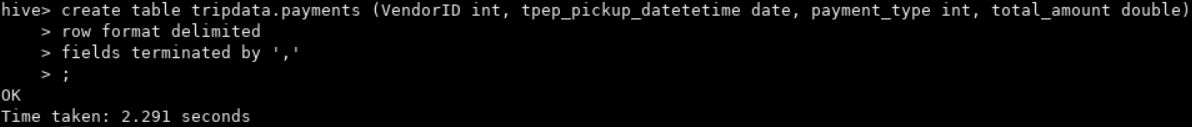
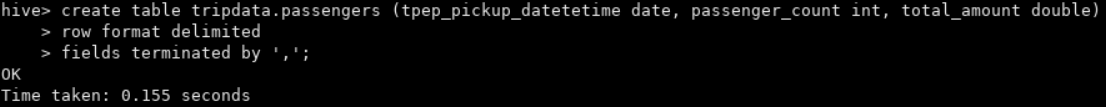
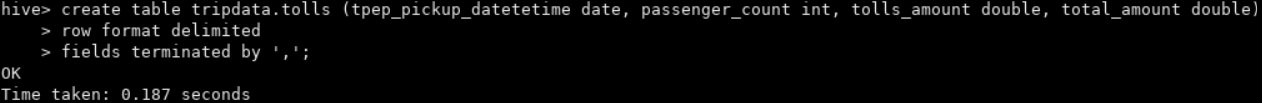
1. En Hive, crear las siguientes tablas (internas) en la base de datos tripdata en hive:
   1. payments:



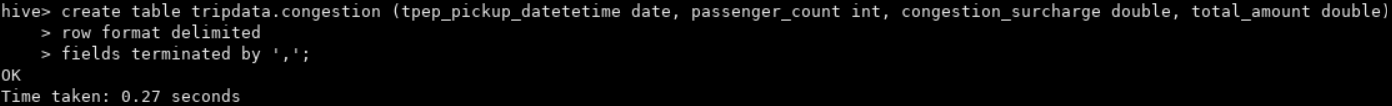
* 1. passengers



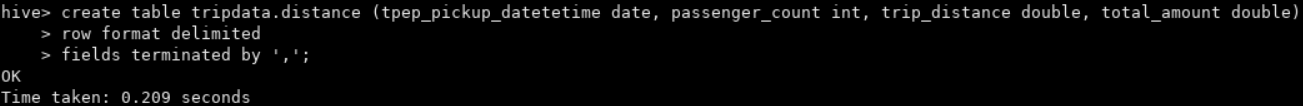
* 1. tolls



* 1. congestion



* 1. distance



1. En Hive, hacer un ‘describe’ de las tablas passengers y distance.

A black background with white text

Description automatically generated

A screen shot of a computer code

Description automatically generated

1. Hacer ingest del file: Yellow\_tripodata\_2021-01.csv <https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2021-01.csv>

Entiendo que el ingest ya lo habíamos hecho con

*wget -P /home/hadoop/landing* [*https://s3.amazonaws.com/nyc-tlc/trip+data/yellow\_tripdata\_2021-01.csv*](https://s3.amazonaws.com/nyc-tlc/trip+data/yellow_tripdata_2021-01.csv)

*hdfs dfs -put /home/hadoop/landing/yellow\_tripdata\_2021-01.csv /ingest*

1. *Opcional*
2. Insertar en la tabla payments (VendorID, tpep\_pickup\_datetetime, payment\_type, total\_amount) Solamente los pagos con tarjeta de crédito

A screenshot of a computer

Description automatically generated

A screen shot of a computer

Description automatically generated

1. Insertar en la tabla passengers (tpep\_pickup\_datetetime, passenger\_count, total\_amount) los registros cuya cantidad de pasajeros sea mayor a 2 y el total del viaje cueste más de 8 dólares.

A screenshot of a computer

Description automatically generated

A screen shot of a black screen

Description automatically generated

1. Insertar en la tabla tolls (tpep\_pickup\_datetetime, passenger\_count, tolls\_amount, total\_amount) los registros que tengan pago de peajes mayores a 0.1 y cantidad de pasajeros mayores a 1.

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

1. Insertar en la tabla congestion (tpep\_pickup\_datetetime, passenger\_count, congestion\_surcharge, total\_amount) los registros que hayan tenido congestión en los viajes en la fecha 2021-01-18

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Insertar en la tabla distance (tpep\_pickup\_datetetime, passenger\_count, trip\_distance, total\_amount) los registros de la fecha 2020-12-31 que hayan tenido solamente un pasajero (passenger\_count = 1) y hayan recorrido más de 15 millas (trip\_distance).

A black screen with white text

Description automatically generated

A black background with white text

Description automatically generated