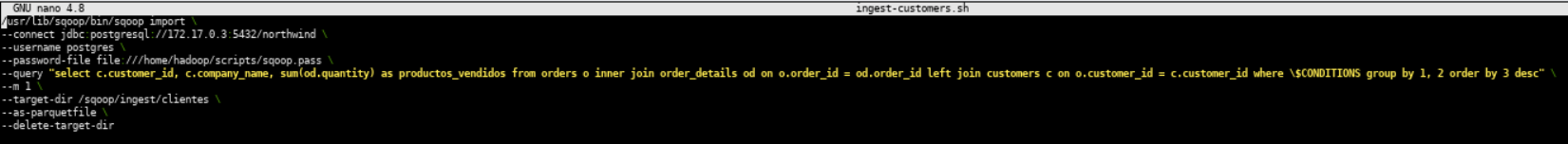
1. Crear una base de datos en Hive llamada northwind\_analytics

A screen shot of a computer

Description automatically generated

1. Crear un script para importar un archivo .parquet de la base northwind que contenga la lista de clientes junto a la cantidad de productos vendidos ordenados de mayor a menor (campos customer\_id, company\_name, productos\_vendidos). Luego ingestar el archivo a HDFS (carpeta /sqoop/ingest/clientes). Pasar la password en un archivo

Creo el script ingest-customers.sh



Creo el archivo sqoop.pass en la misma carpeta que el script haciendo:

**echo -n edvai > sqoop.pass en el directorio /home/hadoop/scripts**

1. Crear un script para importar un archivo .parquet de la base northwind que contenga la lista de órdenes junto a qué empresa realizó cada pedido (campos order\_id, shipped\_date, company\_name, phone). Luego ingestar el archivo a HDFS (carpeta /sqoop/ingest/envíos). Pasar la password en un archivo

Creo el script ingest-shipping.py

A screen shot of a computer

Description automatically generated

1. Crear un script para importar un archivo .parquet de la base northwind que contenga la lista de detalles de órdenes (campos order\_id, unit\_price, quantity, discount). Luego ingestar el archivo a HDFS (carpeta /sqoop/ingest/order\_details). Pasar la password en un archivo

Creo el script ingest-orders.sh

A screenshot of a computer screen

Description automatically generated

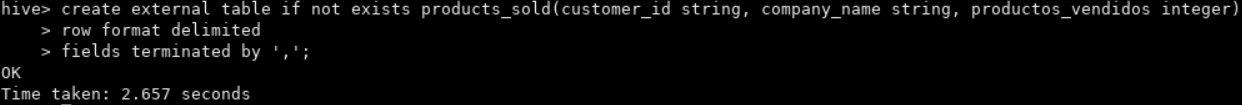
1. Generar un archivo .py que permita mediante Spark insertar en hive en la db northwind\_analytics en la tabla products\_sold, los datos del punto 1, pero solamente aquellas compañías en las que la cantidad de productos vendidos fue mayor al promedio

Creo el script products-sold.py

A screenshot of a computer program

Description automatically generated

Creo la tabla products\_sold



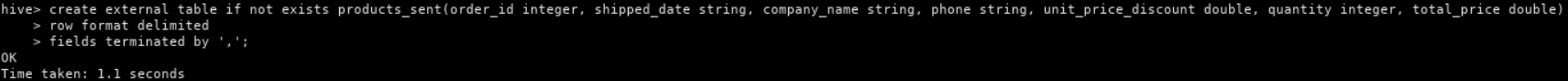
1. Generar un archivo .py que permita mediante Spark insertar en hive en la tabla products\_sent, los datos del punto 2 y 3, de manera tal que se vean las columnas order\_id, shipped\_date, company\_name, phone, unit\_price\_discount (unit\_price with discount), quantity, total\_price (unit\_price\_discount \* quantity). Solo de aquellos pedidos que hayan tenido descuento.

Creo el script products-sent.py

A screenshot of a computer program

Description automatically generated

Creo la tabla products\_sent



Verifico las tablas que he creado en northwind\_analytics

A black screen with white text

Description automatically generated

1. Realizar un proceso automático en Airflow que orqueste los pipelines creados en los puntos anteriores. Crear un grupo para la etapa de ingest y otro para la etapa de process. Correrlo y mostrar una captura de pantalla (del DAG y del resultado en la base de datos)

Creo el DAG (ejercicio9.py)

A screenshot of a computer

Description automatically generated

Lo corro y verifico que termina correctamente

A diagram of a process

Description automatically generated with medium confidence

Verificamos el contenido de products\_sent en Hive

A black screen with white text

Description automatically generated

Verificamos el contenido de products\_sold en Hive

A screen shot of a computer

Description automatically generated