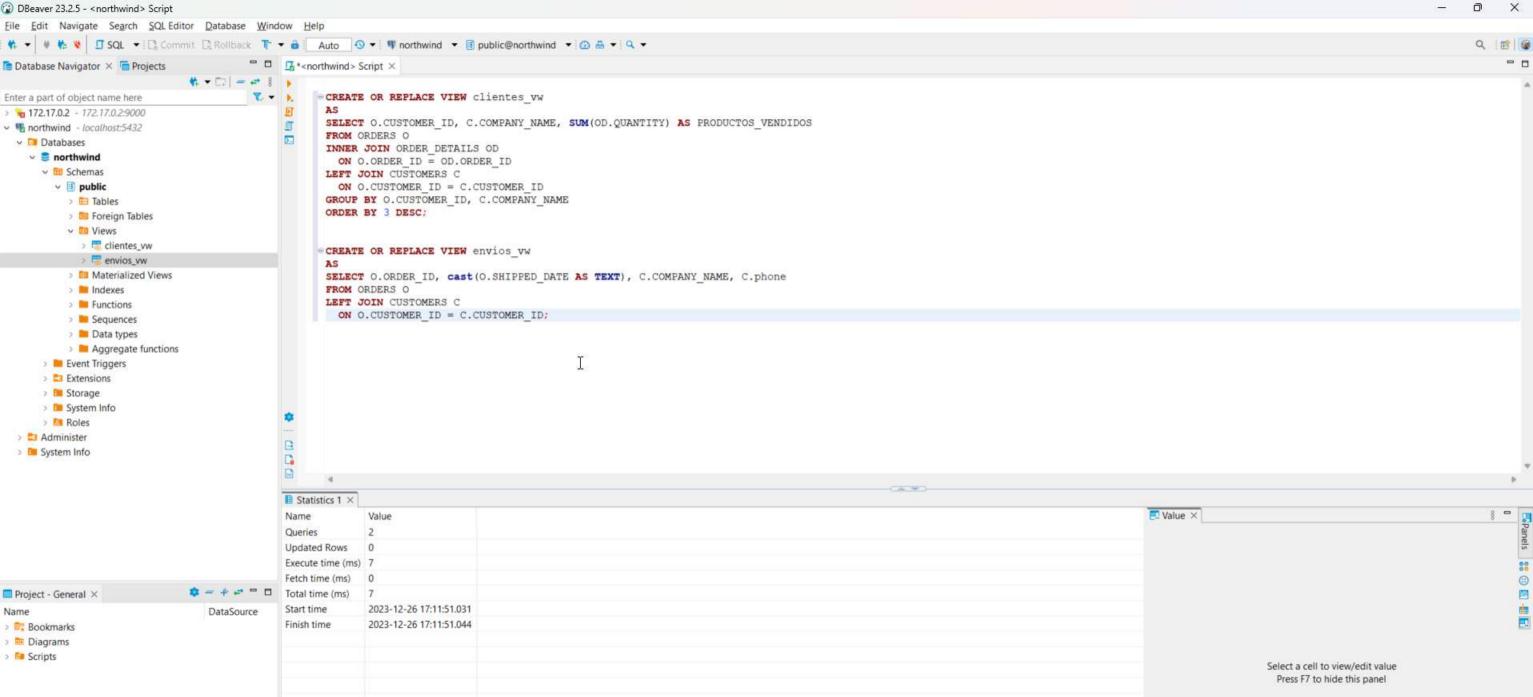
```
hadoop@0e8a607bbfab: /
   hive> CREATE DATABASE IF NOT EXISTS northwind analytics;
OK
Time taken: 0.233 seconds
hive> SHOW DATABASES;
OK
default
northwind analytics
tripdata
tripdata andres leal db
Time taken: 0.027 seconds, Fetched: 4 row(s)
hive>
```



```
o X
                      × № hadoop@0e8a607bbfab: ~/air × № hadoop@0e8a607bbfab: ~/sc ×
 Windows PowerShell
hive> SHOW DATABASES;
default
northwind_analytics
tripdata
tripdata_andres_leal_db
Time taken: 0.469 seconds, Fetched: 4 row(s)
hive> USE northwind analytics;
OK
Time taken: 0.032 seconds
hive> SHOW TABLES;
OK
Time taken: 0.035 seconds
hive> CREATE EXTERNAL TABLE IF NOT EXISTS northwind_analytics.products_sent(ORDER_ID_INT, SHIPPED_DATE TIMESTAMP, COMPANY_NAME STRING, phone STRING, UNIT_PR
ICE DISCOUNT DOUBLE, QUANTITY DOUBLE, TOTAL PRICE DOUBLE)
   > COMMENT 'products sent ONLY LISTS ORDERS THAT HAD A DISCOUNT'
   > ROW FORMAT DELIMITED
   > FIELDS TERMINATED BY ',';
OK
Time taken: 1.005 seconds
hive> CREATE EXTERNAL TABLE IF NOT EXISTS northwind_analytics.products_sold(CUSTOMER_ID STRING, COMPANY_NAME STRING, PRODUCTOS_VENDIDOS INT)
   > COMMENT 'products sold CONTAINS THE LIST OF CLIENTS WITH A NUM OF PRODUCTS SOLD ABOVE THE AVG'
   > ROW FORMAT DELIMITED
   > FIELDS TERMINATED BY ',';
OK
Time taken: 0.085 seconds
hive> SHOW TABLES;
products_sent
products_sold
Time taken: 0.036 seconds, Fetched: 2 row(s)
hive>
```

```
hadoop@0e8a607bbfab:~/scripts$ ls -lh
total 72K
-rw-rw-r-- 1 hadoop hadoop 9.1K Dec 26 01:22 QueryResult.java
-rw-rw-r-- 1 hadoop hadoop  15K Dec 26 01:41 codegen clientes vw.java
-rw-rw-r-- 1 hadoop hadoop 16K Dec 26 01:38 codegen order details.java
-rw-rw-r-- 1 hadoop hadoop 670 Feb 28 2022 derby.log
-rwxrwxr-x 1 hadoop hadoop 258 Feb 28 2022 ingest.sh
-rwxrwxrwx 1 hadoop hadoop 2.3K Dec 25 15:17 ingest andresleal.sh
drwxr-xr-x 2 hadoop hadoop 4.0K Feb 9 2022 spark-warehouse
-rwxrwxrwx 1 hadoop hadoop 1.1K May 9 2022 start-services.sh
-rwxrwxrwx 1 hadoop hadoop 1.1K May 9 2022 transformation.py
-rwxrwxrwx 1 hadoop hadoop 1.3K Dec 25 15:51 transformation andresleal.py
hadoop@0e8a607bbfab:~/scripts$ nano ingest order details.sh
hadoop@0e8a607bbfab:~/scripts$ nano ingest clientes.sh
hadoop@0e8a607bbfab:~/scripts$ nano ingest envios.sh
hadoop@0e8a607bbfab:~/scripts$ ls -lh
total 84K
-rw-rw-r-- 1 hadoop hadoop 9.1K Dec 26 01:22 QueryResult.java
-rw-rw-r-- 1 hadoop hadoop 15K Dec 26 01:41 codegen clientes vw.java
-rw-rw-r-- 1 hadoop hadoop 16K Dec 26 01:38 codegen order details.java
-rw-rw-r-- 1 hadoop hadoop 670 Feb 28 2022 derby.log
-rwxrwxr-x 1 hadoop hadoop 258 Feb 28 2022 ingest.sh
-rwxrwxrwx 1 hadoop hadoop 2.3K Dec 25 15:17 ingest andresleal.sh
-rw-rw-r-- 1 hadoop hadoop 217 Dec 26 01:44 ingest clientes.sh
-rw-rw-r-- 1 hadoop hadoop 213 Dec 26 01:45 ingest envios.sh
-rw-rw-r-- 1 hadoop hadoop 279 Dec 26 01:44 ingest order details.sh
drwxr-xr-x 2 hadoop hadoop 4.0K Feb 9 2022 spark-warehouse
-rwxrwxrwx 1 hadoop hadoop 1.1K May 9 2022 start-services.sh
-rwxrwxrwx 1 hadoop hadoop 1.1K May 9 2022 transformation.py
-rwxrwxrwx 1 hadoop hadoop 1.3K Dec 25 15:51 transformation andresleal.pv
hadoop@0e8a607bbfab:~/scripts$ chmod 777 ingest order details.sh ingest clientes.sh ingest envios.sh
hadoop@0e8a607bbfab:~/scripts$ ls -lh
total 84K
-rw-rw-r-- 1 hadoop hadoop 9.1K Dec 26 01:22 QueryResult.java
-rw-rw-r-- 1 hadoop hadoop 15K Dec 26 01:41 codegen clientes vw.java
-rw-rw-r-- 1 hadoop hadoop 16K Dec 26 01:38 codegen order details.java
-rw-rw-r-- 1 hadoop hadoop 670 Feb 28 2022 derby.log
-rwxrwxr-x 1 hadoop hadoop 258 Feb 28 2022 ingest.sh
-rwxrwxrwx 1 hadoop hadoop 2.3K Dec 25 15:17 ingest andresleal.sh
-rwxrwxrwx 1 hadoop hadoop 217 Dec 26 01:44 ingest clientes.sh
-rwxrwxrwx 1 hadoop hadoop 213 Dec 26 01:45 ingest envios.sh
-rwxrwxrwx 1 hadoop hadoop 279 Dec 26 01:44 ingest order details.sh
drwxr-xr-x 2 hadoop hadoop 4.0K Feb 9 2022 spark-warehouse
-rwxrwxrwx 1 hadoop hadoop 1.1K May 9 2022 start-services.sh
```

```
PS C:\Users\DELL> docker exec -it edvai hadoop bash
root@0e8a607bbfab:/# su hadoop
hadoop@0e8a607bbfab:/$
hadoop@0e8a607bbfab:/$ cd home/hadoop/scripts/
hadoop@0e8a607bbfab:~/scripts$ ls -lh
total 104K
-rw-rw-r-- 1 hadoop hadoop 9.1K Dec 26 01:22 QueryResult.java
-rw-rw-r-- 1 hadoop hadoop 15K Dec 26 01:41 codegen clientes vw.java
-rw-rw-r-- 1 hadoop hadoop 17K Dec 26 01:48 codegen_envios_vw.java
-rw-rw-r-- 1 hadoop hadoop 670 Feb 28 2022 derby.log
-rwxrwxr-x 1 hadoop hadoop 258 Feb 28 2022 ingest.sh
-rwxrwxrwx 1 hadoop hadoop 2.3K Dec 25 15:17 ingest andresleal.sh
-rwxrwxrwx 1 hadoop hadoop 217 Dec 26 01:44 ingest clientes.sh
-rwxrwxrwx 1 hadoop hadoop 213 Dec 26 01:45 ingest envios.sh
rwxrwxrwx 1 hadoop hadoop 279 Dec 26 01:44 ingest_order_details.sh
drwxr-xr-x 2 hadoop hadoop 4.0K Feb 9 2022 spark-warehouse
-rwxrwxrwx 1 hadoop hadoop 1.1K May 9 2022 start-services.sh
-rwxrwxrwx 1 hadoop hadoop 1.1K May 9 2022 transformation.py
-rwxrwxrwx 1 hadoop hadoop 1.3K Dec 25 15:51 transformation_andresleal.py
hadoop@0e8a607bbfab:~/scripts$ nano transform_load_clientes.py
hadoop@0e8a607bbfab:~/scripts$ nano transform_envios_order_details.py
hadoop@0e8a607bbfab:~/scripts$ cat transform envios order details.py
from pyspark.context import SparkContext
from pyspark.sql.session import SparkSession
sc = SparkContext('local')
spark = SparkSession(sc)
from pyspark.sql import HiveContext
hc = HiveContext(sc)
# CARGAMOS LOS DF, Y CREAMOS VISTAS TEMPORALES EN SPARK PARA SU MANIPULACION CON SQL
envios_df = spark.read.parquet("/sqoop/ingest/envios/*.parquet")
order_details_df = spark.read.parquet("/sqoop/ingest/order_details/*.parquet")
envios_df.createOrReplaceTempView("v_envios")
order details df.createOrReplaceTempView("v od")
# SQL STMNT PARA REALIZAR LOS CALCULOS REQUERIDOS
select orders = """
select
v envios.*,
unit_price * (1-discount) as unit_price_discount,
quantity,
(unit price * (1-discount) ) * quantity as total price
from v envios
left join v od
on v envios.order id = v od.order id
where discount > 0
# INSERTAMOS LA DATA PROCESADA EN LA TABLA DESEADA
hive_insert_stmnt = f'insert into northwind_analytics.products_sent {select_orders};'
hc.sql(hive insert stmnt)
hadoop@0e8a607bbfab:~/scripts$
hadoop@0e8a607bbfab:~/scripts$ ls -lh
total 112K
-rw-rw-r-- 1 hadoop hadoop 9.1K Dec 26 01:22 QueryResult.java
-rw-rw-r-- 1 hadoop hadoop 15K Dec 26 01:41 codegen_clientes_vw.java
-rw-rw-r-- 1 hadoop hadoop 17K Dec 26 01:48 codegen_envios_vw.java
-rw-rw-r-- 1 hadoop hadoop 16K Dec 26 01:38 codegen order details.java
-rw-rw-r-- 1 hadoop hadoop 670 Feb 28 2022 derby.log
-rwxrwxr-x 1 hadoop hadoop 258 Feb 28 2022 ingest.sh
-rwxrwxrwx 1 hadoop hadoop 2.3K Dec 25 15:17 ingest andresleal.sh
 rwxrwxrwx 1 hadoop hadoop 217 Dec 26 01:44 ingest clientes.s
-rwxrwxrwx 1 hadoop hadoop 213 Dec 26 01:45 ingest envios.sh
-rwxrwxrwx 1 hadoop hadoop 279 Dec 26 01:44 ingest order details.sh
drwxr-xr-x 2 hadoop hadoop 4.0K Feb 9 2022 spark-warehouse
-rwxrwxrwx 1 hadoop hadoop 1.1K May 9 2022 start-services.sh
-rw-rw-r-- 1 hadoop hadoop 984 Dec 26 15:21 transform envios order details.py
-rw-rw-r-- 1 hadoop hadoop  780 Dec 26 15:20 transform load clientes.py
rwxrwxrwx 1 hadoop hadoop 1.1K May 9 2022 transformation.py
rwxrwxrwx 1 hadoop hadoop 1.3K Dec 25 15:51 transformation andresleal.py
nadoop@0e8a607bbfab:~/scripts$ chmod 777 transform envios order details.py transform load clientes.py
hadoop@0e8a607bbfab:~/scripts$ Is -Ih
total 112K
-rw-rw-r-- 1 hadoop hadoop 9.1K Dec 26 01:22 QueryResult.java
-rw-rw-r-- 1 hadoop hadoop 15K Dec 26 01:41 codegen clientes vw.java
-rw-rw-r-- 1 hadoop hadoop  17K Dec 26 01:48 codegen envios vw.java
-rw-rw-r-- 1 hadoop hadoop 16K Dec 26 01:38 codegen order details.java
-rw-rw-r-- 1 hadoop hadoop 670 Feb 28 2022 derby.log
-rwxrwxr-x 1 hadoop hadoop 258 Feb 28 2022 ingest.sh
rwxrwxrwx 1 hadoop hadoop 2.3K Dec 25 15:17 ingest andresleal.sh
-rwxrwxrwx 1 hadoop hadoop 217 Dec 26 01:44 ingest clientes.sh
-rwxrwxrwx 1 hadoop hadoop 213 Dec 26 01:45 ingest envios.sh
-rwxrwxrwx 1 hadoop hadoop 279 Dec 26 01:44 ingest order details.sh
drwxr-xr-x 2 hadoop hadoop 4.0K Feb 9 2022 spark-warehouse
-rwxrwxrwx 1 hadoop hadoop 1.1K May 9 2022 start-services.sh
-rwxrwxrwx 1 hadoop hadoop 984 Dec 26 15:21 transform_envios_order_details.py
-rwxrwxrwx 1 hadoop hadoop 780 Dec 26 15:20 transform load clientes.py
-rwxrwxrwx 1 hadoop hadoop 1.1K May 9 2022 transformation.py
rwxrwxrwx 1 hadoop hadoop 1.3K Dec 25 15:51 transformation andresleal.py-
hadoop@0e8a607bbfab:~/scripts$|
```

💹 hadoop@0e8a607bbfab: ~/sc 🛛 💹 hadoop@0e8a607bbfab: ~/air 🗶 🕂 🔻

```
PS C:\Users\DELL> docker exec -it edvai hadoop bash
root@0e8a607bbfab:/# su hadoop
hadoop@0e8a607bbfab:/$
hadoop@0e8a607bbfab:/$ cd home/hadoop/airflow/dags/
hadoop@0e8a607bbfab:~/airflow/dags$ nano ingest transform ejercicio9 andresleal.py
hadoop@0e8a607bbfab:~/airflow/dags$ Is -Ih
total 20K
drwxrwxr-x 1 hadoop hadoop 4.0K Dec 25 15:47 pycache
-rw-rw-r-- 1 hadoop hadoop 1.1K May 1 2022 example-DAG.py
-rw-rw-r-- 1 hadoop hadoop 1.0K May 5 2022 ingest-transform.py
-rw-rw-r-- 1 hadoop hadoop 1.2K Dec 25 15:47 ingest transform andresleal.py
-rw-rw-r-- 1 hadoop hadoop 2.3K Dec 26 15:23 ingest transform ejercicio9 andresleal.py
hadoop@0e8a607bbfab:~/airflow/dags$
```

