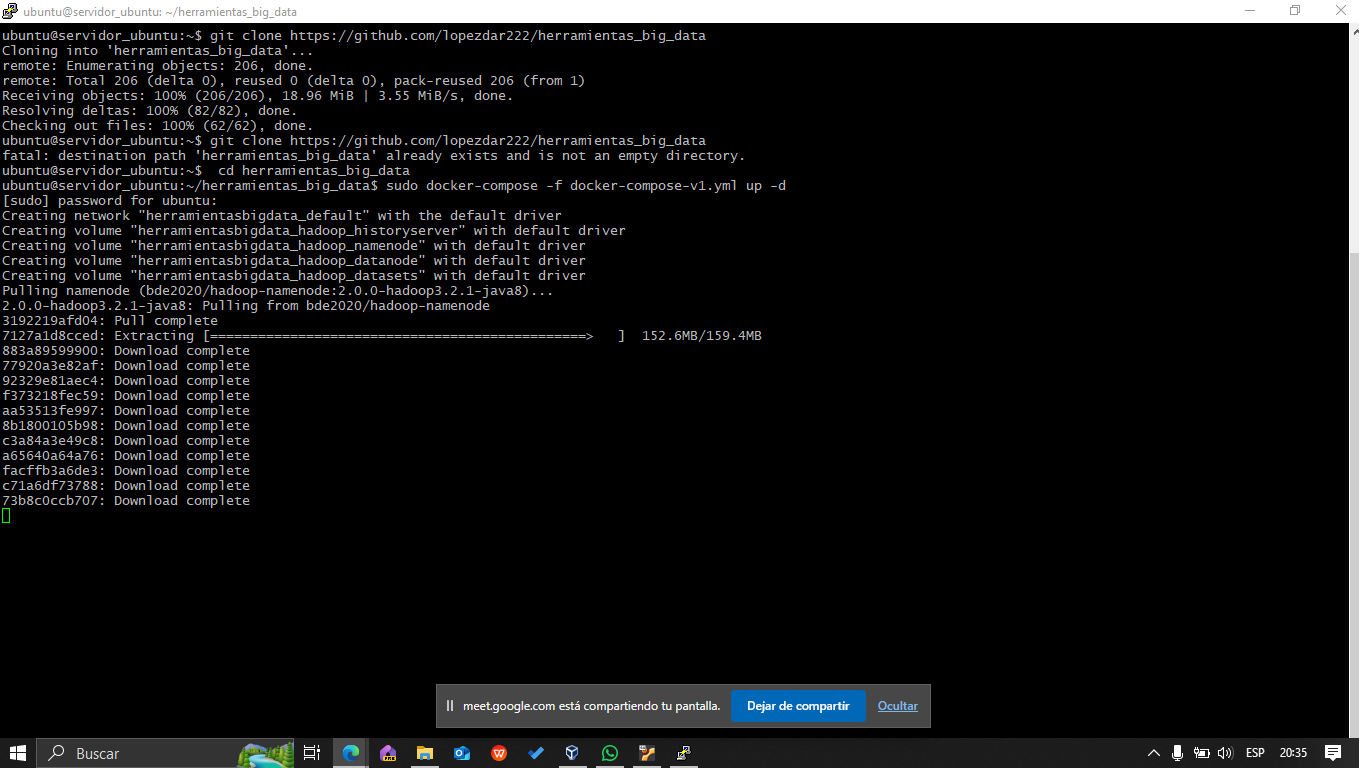
1. **HDFS: copiar los archivos csv del Datasets al contenedor namenode y al HDFS**

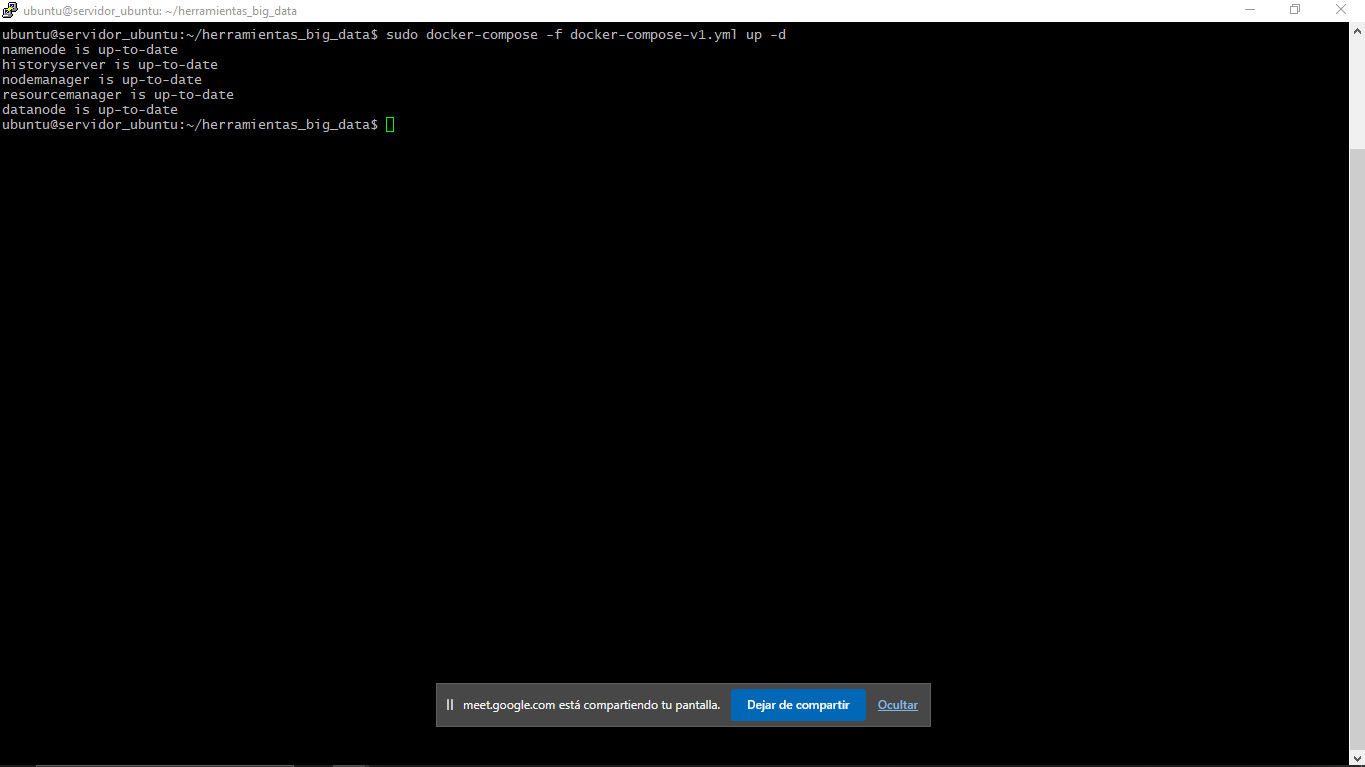
Paso 1: clonación de /herramientas\_big\_data



Paso 2: cd herramientas\_big\_data

Paso 3: iniciación del ecosistema de hadoop mediante docker-compose

sudo docker-compose -f docker-compose-v1.yml up -d



Paso 4: ubicarse en el contenedor nanenode

sudo docker exec -it namenode bash

Paso 5: crear el directorio Datasets en el contenedor

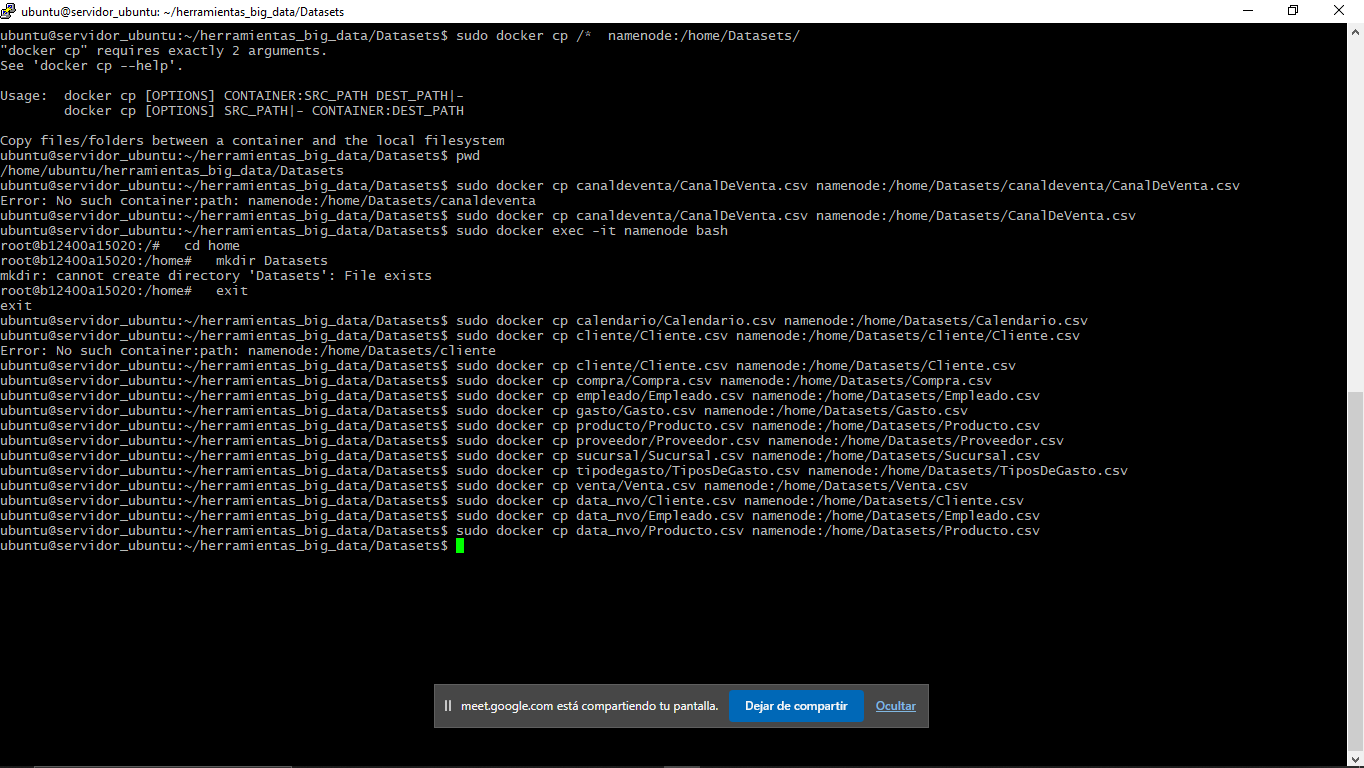
cd home

mkdir Datasets

Exit

Paso 6: copiar todos los archivos desde /herramientas\_big\_data al contenedor namenode

sudo docker cp <path><archivo> namenode:/home/Datasets/<archivo>



Paso 7: Ubicarse en el contenedor "namenode"

sudo docker exec -it namenode bash

Paso 8: Crear un directorio en HDFS llamado "/data"

hdfs dfs -mkdir -p /data

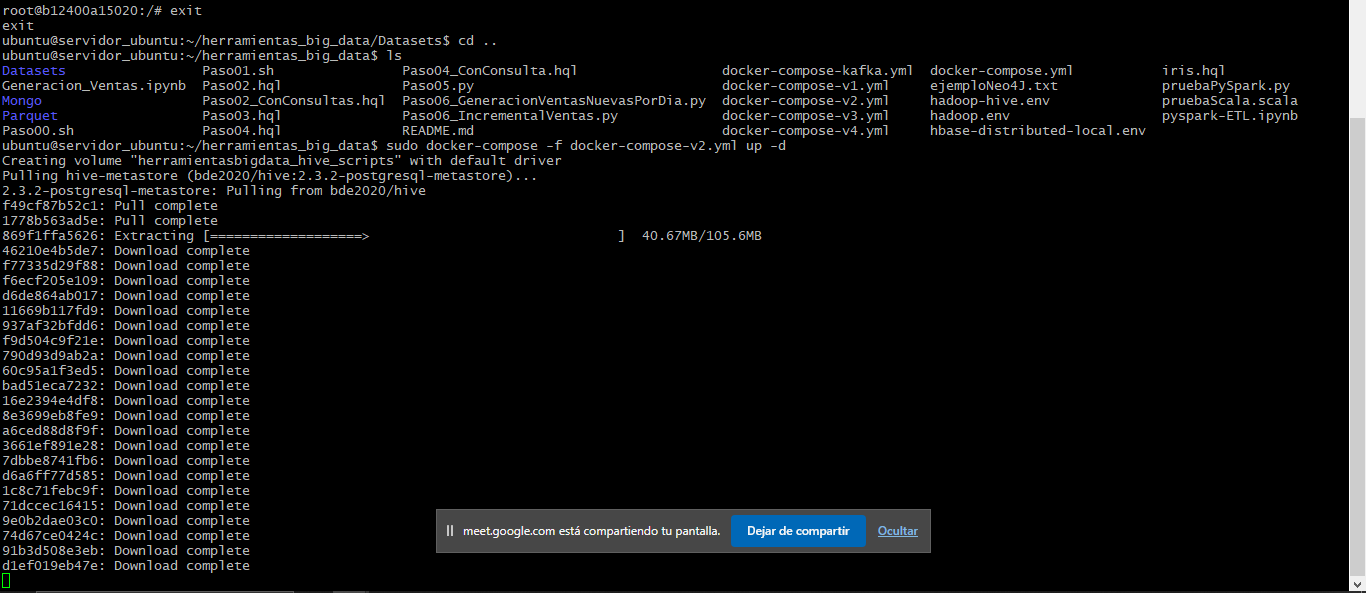
Paso 9: Copiar los archivos csv provistos a HDFS:

hdfs dfs -put /home/Datasets/\* /data

1. Hive

Paso 1: Se utiliza el entorno docker-compose-v2.yml desde el directorio de /herramientas\_big\_data/Datasets

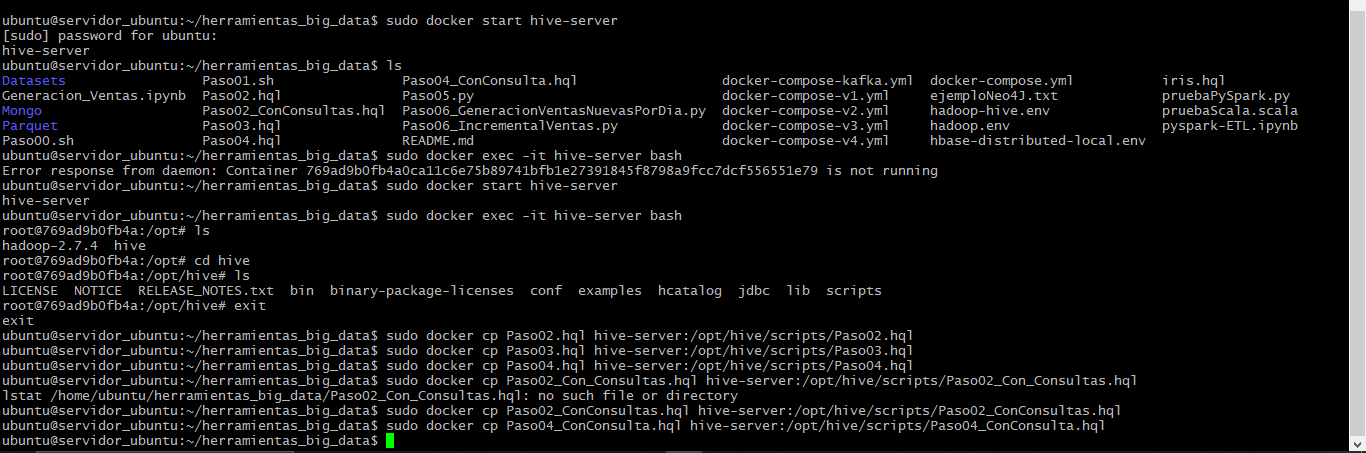
sudo docker-compose -f docker-compose-v2.yml up -d



Paso 2: Crear tablas en Hive, a partir de los csv ingestados en HDFS

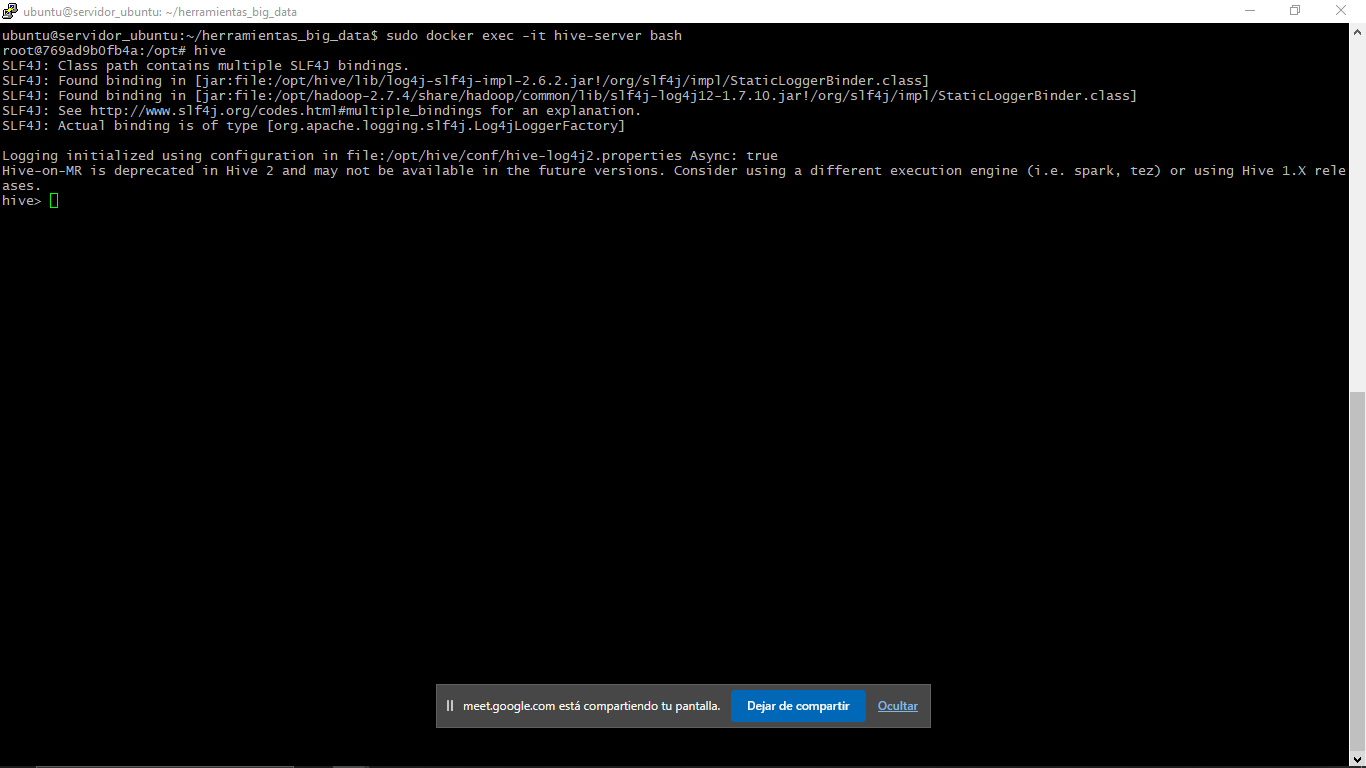
Para ello, primero es necesario copiar los scripts de hql desde el directorio de /herramientas\_big\_data al contenedor hive-server utilizando la siguiente sentencia:

sudo docker cp <path><archivo> hive-server:/opt/<archivo>



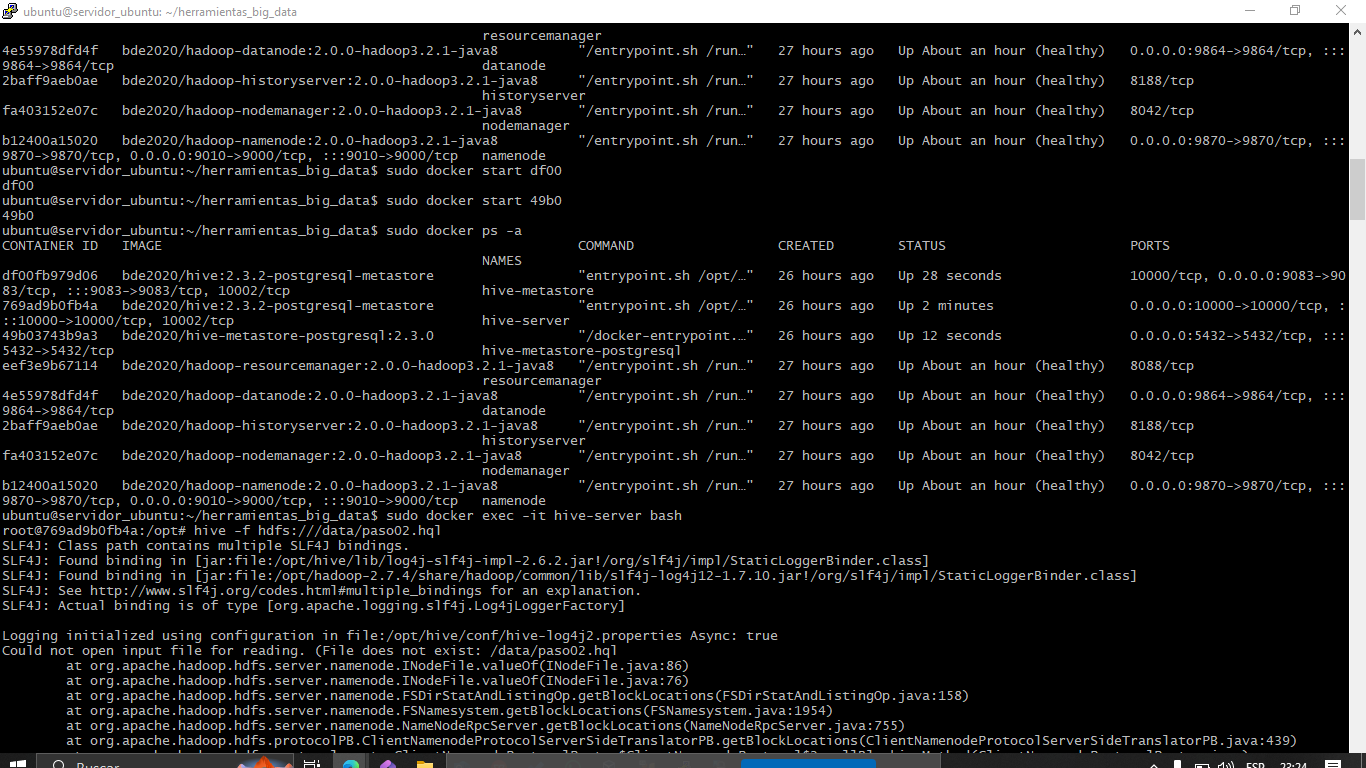
Para esto, se ubica el contenedor correspondiente al servidor de Hive

sudo docker exec -it hive-server bash



Ahora se crean las tablas en Hive utilizando la siguiente sentencia:

hive -f <script.hql>



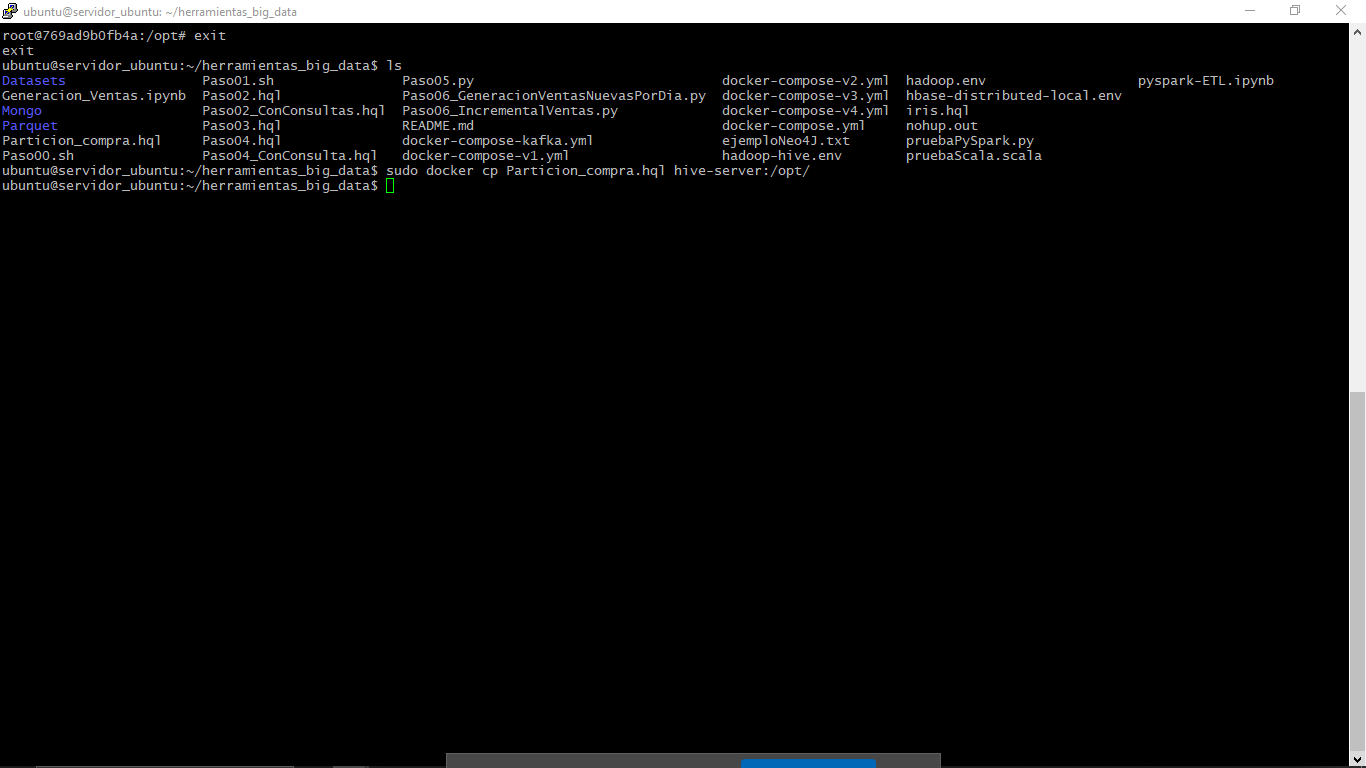
1. Formato de almacenamiento

Se creó un archivo llamado Particion\_compra.hql

Paso 1: copiar el archivo desde mi computadora a la maquina virtual en herramientas\_big\_data

Paso 2: copia el archivo en el contenedor de hive-server

Sudo docker cp Particion\_compra.hql hive-server:/opt/

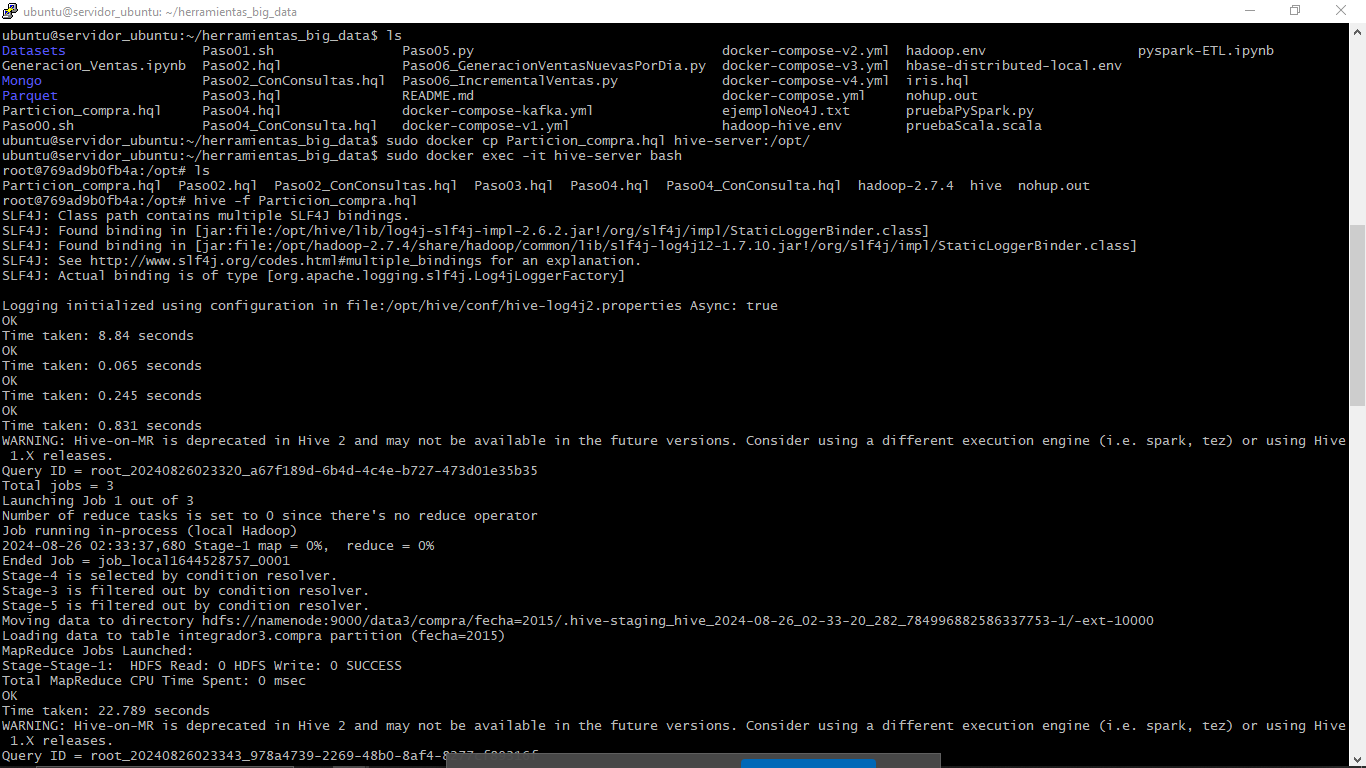


Paso 3: entro en el contenedor de hive-server

Sudo docker exec -it hive-server bash

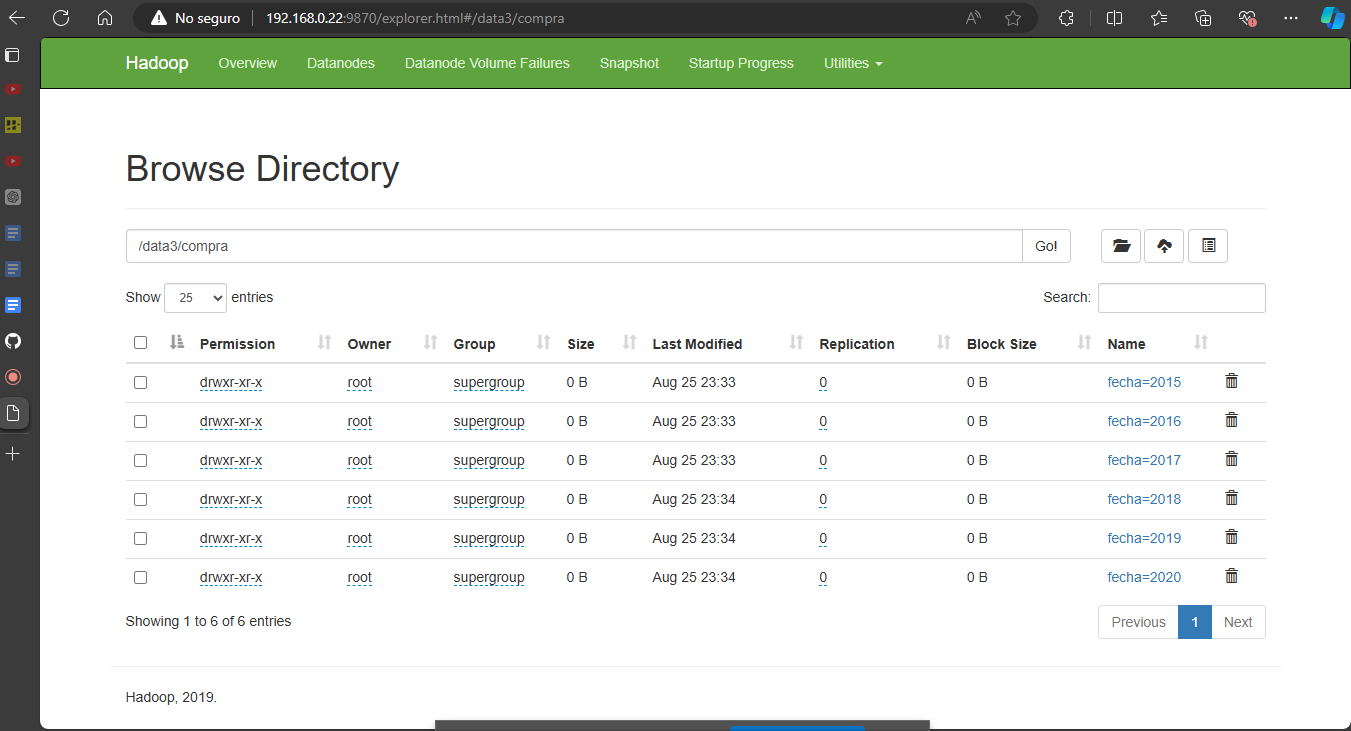
Paso 4: ejecuto el script

Hive -f Particion\_compra.hql



1. SQL

Observación de la partición de Particion\_compra.hql

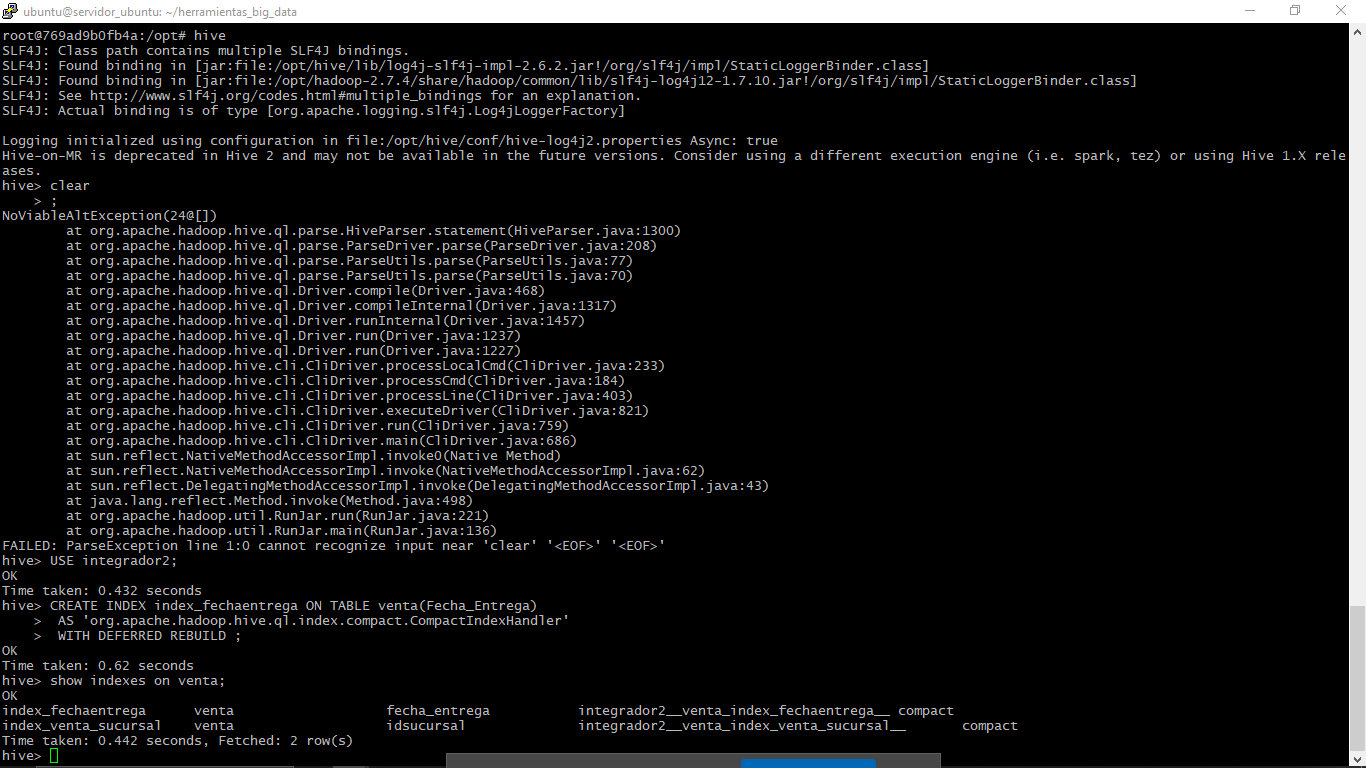


Inicio el motor de hive

Me conecto a la base de datos

USE integrador 2;

Y creo el indice



Listamos la descripción de la tabla completa:

• describe formatted venta;

