Done By: Esraa Ahmed Fouad Omar # ID: 20399123

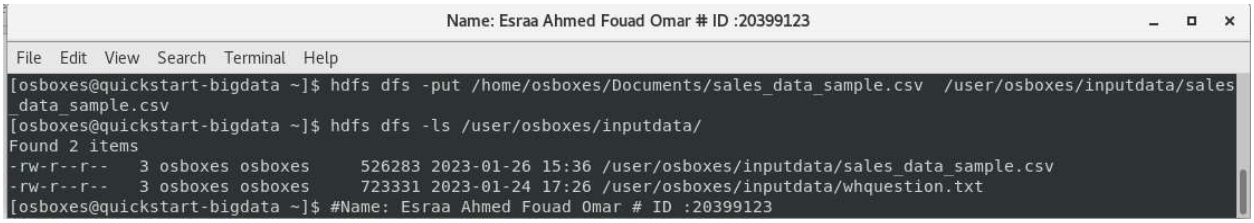
Assignment #: 2

Done by: Esraa Khaled Ahmed Fouad Omar # ID: 20399123

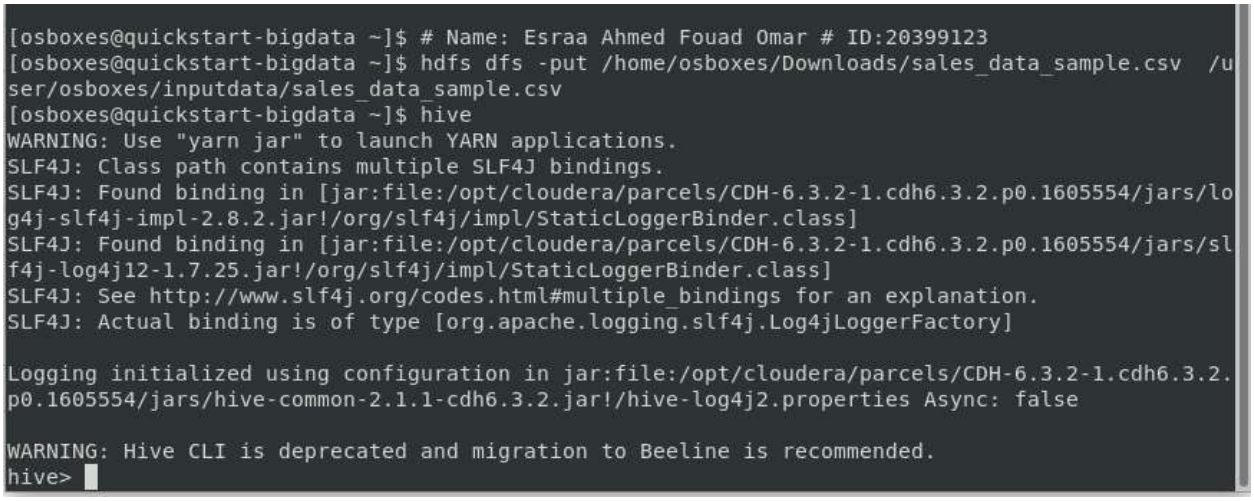
January 23, 2023

Part 2

Copy sales_data_samples.csv from local virtual machine to HDFS. As you can see, we copied the file successfully.



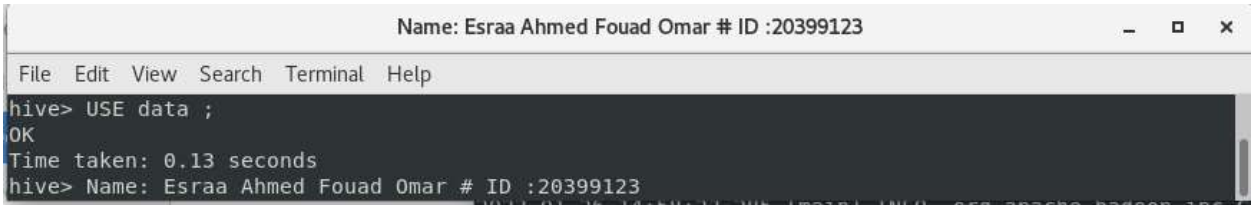
Open the terminal and write hive in it to be in hive mode. As you can see, we got in the hive mode successfully.



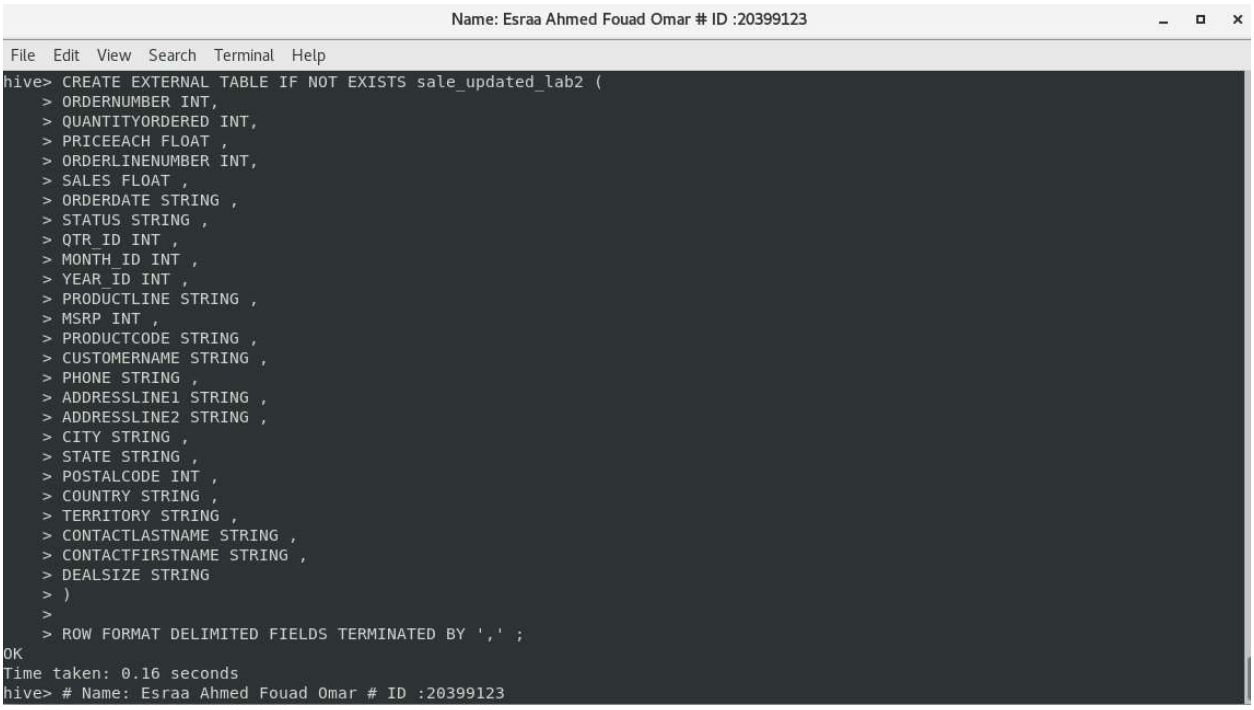
Create Database and call it data. As you can see, we create database successfully.

Done By: Esraa Ahmed Fouad Omar # ID: 20399123

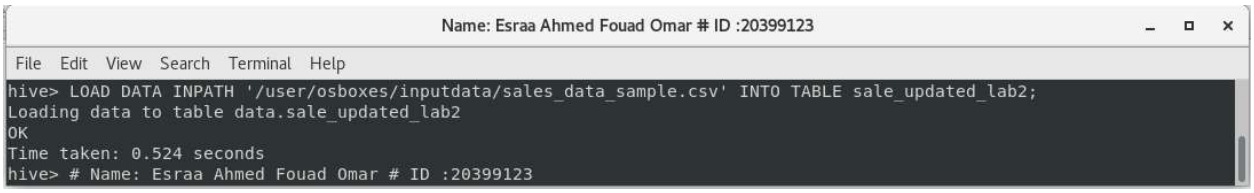
Use database which we were created from previous step. As you can see, everything is ok.



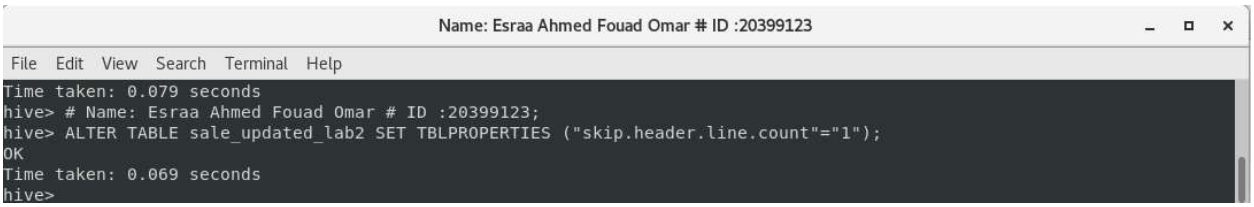
Create table and insert the name and data type of columns into it as shown below.



Load data(sales_data_sample.csv) into the sale_updated_lab2 table as shown below.

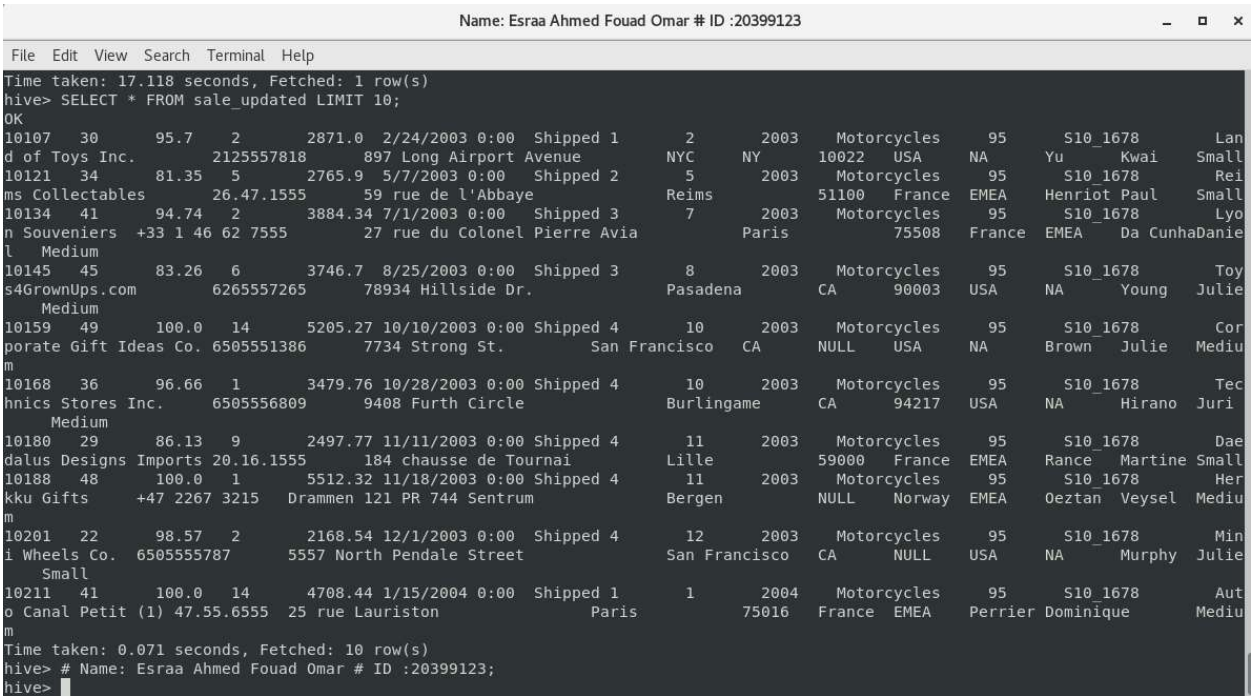


Skip/remove the header from the table as shown below



Done By: Esraa Ahmed Fouad Omar # ID: 20399123

Let’s see sample of our data.



Done By: Esraa Ahmed Fouad Omar # ID: 20399123

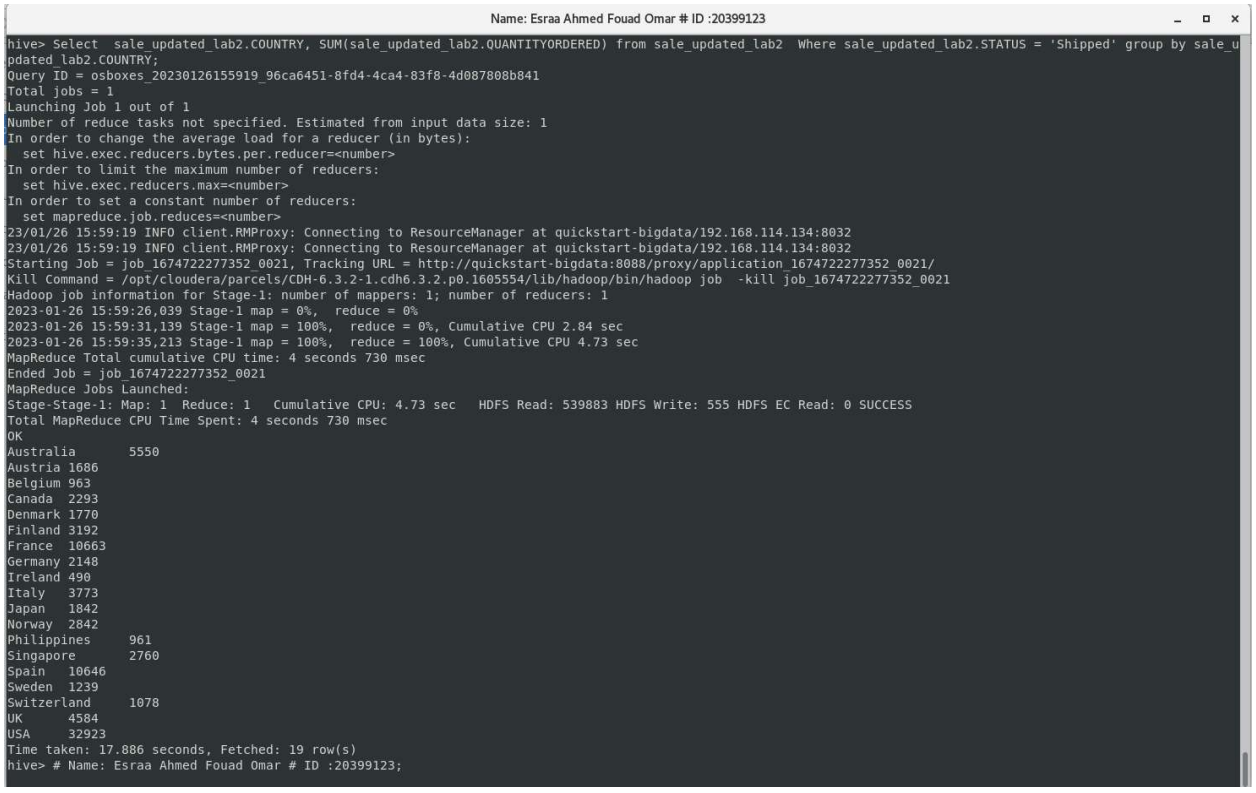
1. Number of orders per country

Name: Esraa Ahmed Fouad Omar # ID :20399123

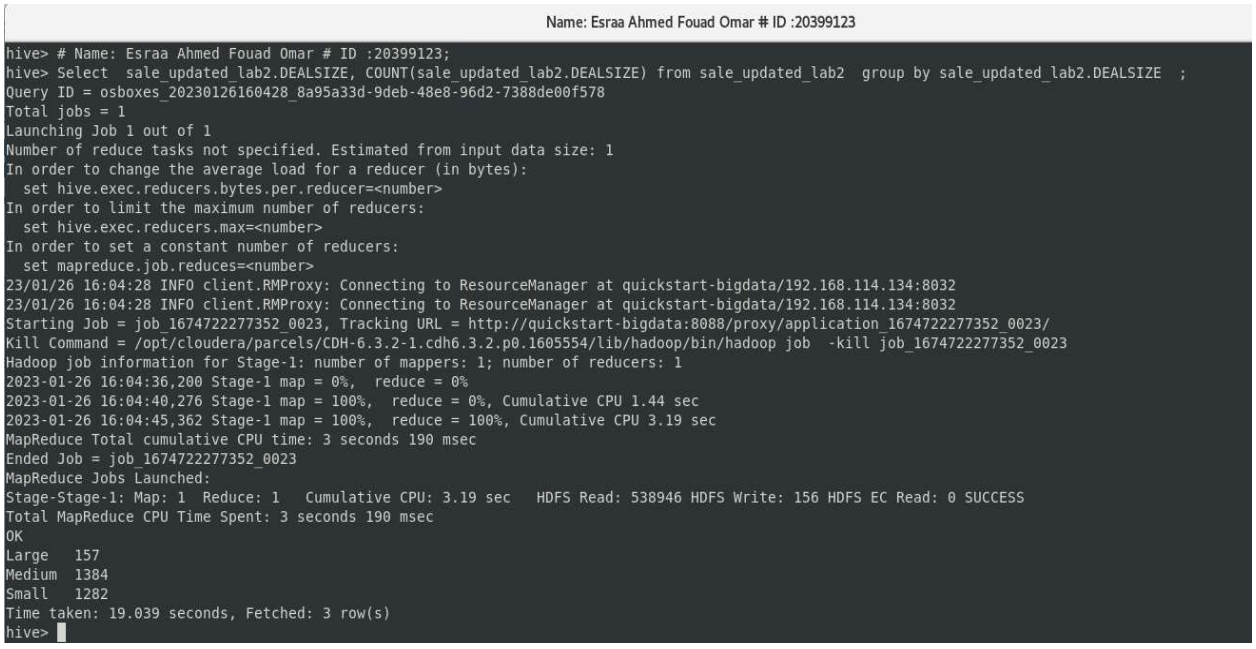
```
hive> # Name: Esraa Ahmed Fouad Omar # ID :20399123;
hive> SELECT sale_updated_lab2.COUNTRY , COUNT(sale_updated_lab2.ORDERLINENUMBER) count FROM sale_updated_lab2
    > GROUP BY sale_updated_lab2.COUNTRY;
Query ID = osboxes_20230126160226_c6931d9c-d852-4900-b862-4aebfdfdc43e
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/01/26 16:02:26 INFO client.RMProxy: Connecting to ResourceManager at quickstart-bigdata/192.168.114.134:8032
23/01/26 16:02:26 INFO client.RMProxy: Connecting to ResourceManager at quickstart-bigdata/192.168.114.134:8032
Starting Job = job_1674722277352_0022, Tracking URL = http://quickstart-bigdata:8088/proxy/application_1674722277352_0022/
Kill Command = /opt/cloudera/parcels/CDH-6.3.2-1.cdh6.3.2.p0.1605554/lib/hadoop/bin/hadoop job  -kill job_1674722277352_0022
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-01-26 16:02:35,094 Stage-1 map = 0%,  reduce = 0%
2023-01-26 16:02:39,193 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.42 sec
2023-01-26 16:02:44,313 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.19 sec
MapReduce Total cumulative CPU time: 3 seconds 190 msec
Ended Job = job_1674722277352_0022
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.19 sec   HDFS Read: 538949 HDFS Write: 524 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 190 msec
OK
Australia       185
Austria 55
Belgium 33
Canada  70
Denmark 63
Finland 92
France  314
Germany 62
Ireland 16
Italy   113
Japan   52
Norway  85
Philippines     26
Singapore       79
Spain   342
Sweden  57
Switzerland     31
UK      144
USA     1004
Time taken: 19.374 seconds, Fetched: 19 row(s)
hive>
```

Done By: Esraa Ahmed Fouad Omar # ID: 20399123

2. Total amount for the orders that are shipped (status=shipped) per country



3. Total how many Small, Medium and Large deal have been made.



Done By: Esraa Ahmed Fouad Omar # ID: 20399123

4. create separate partition for order STATUS

- create table with the same columns name and data type and partition it using status column as shown below.

```
                                                    Name: Esraa Ahmed Fouad Omar # ID :20399123
hive> # Name: Esraa Ahmed Fouad Omar # ID :20399123;
hive> CREATE TABLE IF NOT EXISTS partitioned_sale_updated_lab2
    > (
    > ORDERNUMBER INT,
    > QUANTITYORDERED INT,
    > PRICEEACH FLOAT ,
    > ORDERLINENUMBER INT,
    > SALES FLOAT ,
    > ORDERDATE STRING ,
    > QTR_ID INT ,
    > MONTH_ID INT ,
    > YEAR_ID INT ,
    > PRODUCTLINE STRING ,
    > MSRP INT ,
    > PRODUCTCODE STRING ,
    > CUSTOMERNAME STRING ,
    > PHONE STRING ,
    > ADDRESSLINE1 STRING ,
    > ADDRESSLINE2 STRING ,
    > CITY STRING ,
    > STATE STRING ,
    > POSTALCODE INT ,
    > COUNTRY STRING ,
    > TERRITORY STRING ,
    > CONTACTLASTNAME STRING ,
    > CONTACTFIRSTNAME STRING ,
    > DEALSIZE STRING )
    > partitioned by (STATUS STRING);
OK
Time taken: 0.096 seconds
hive>
```

- Change your configuration from strict mode to non-strict mode. As you can see, we changed it successfully

```
                                                    Name: Esraa Ahmed Fouad Omar # ID :20399123
hive> set hive.exec.dynamic.partition.mode=nonstrict;
hive> # Name: Esraa Ahmed Fouad Omar # ID :20399123;
hive>
```

Done By: Esraa Ahmed Fouad Omar # ID: 20399123

- Overwrite into the previous created table as shown below.

- Describe your table.