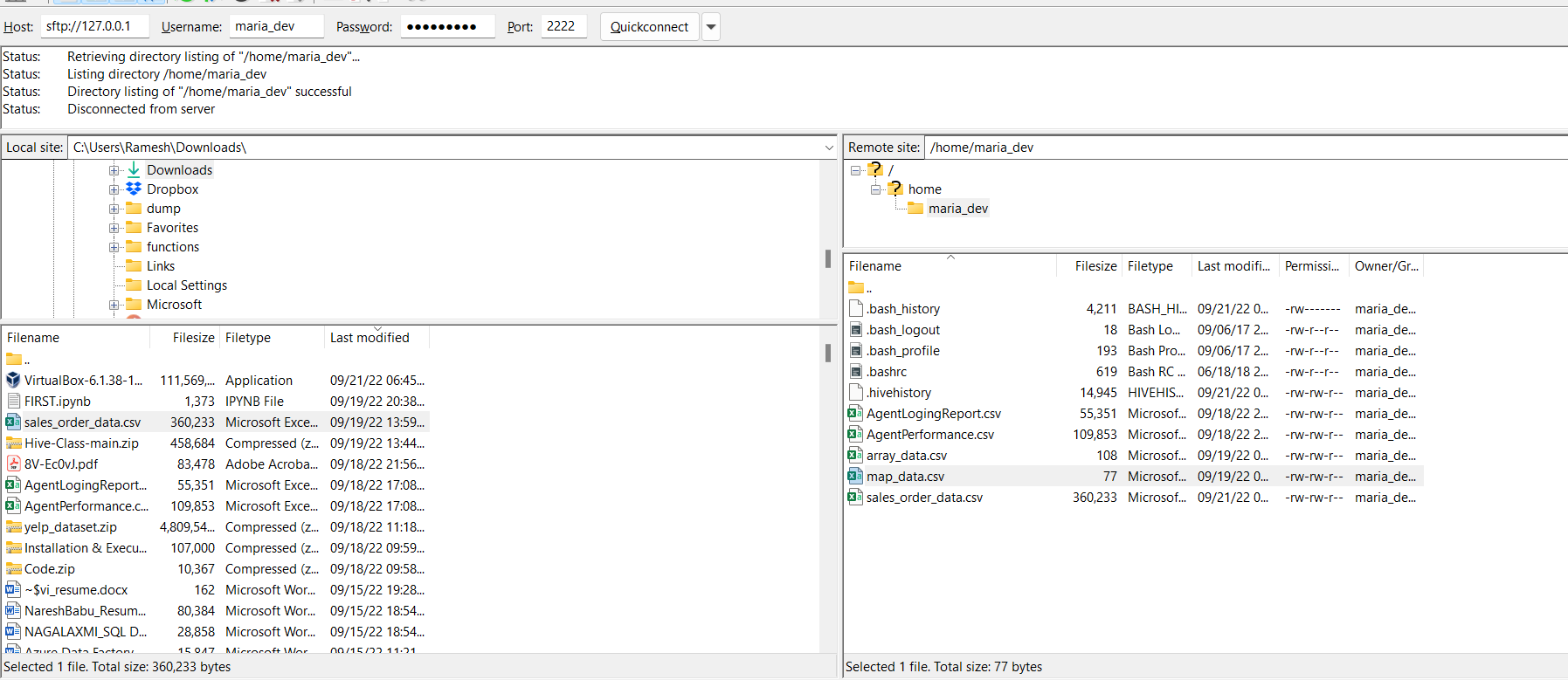
**Step 1**: Download vehicle sales data from below link.

https://github.com/shashank-mishra219/Hive-Class/blob/main/sales\_order\_data.csv

**Step 2**: Copied the data set from laptop to Linux machine using FileZilla.

Path: /home/maria\_dev



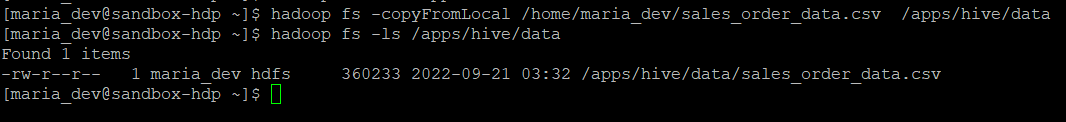
**Step 3**: Created directory to store the Data set into Hadoop.

Hadoop fs -mkdir /apps/hive/data



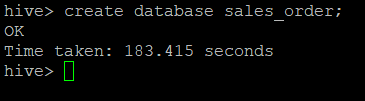
S**tep 4**: Copied dataset from local to Hadoop directory.

hadoop fs -copyFromLocal /home/maria\_dev/ sales\_order\_data.csv /apps/hive/data

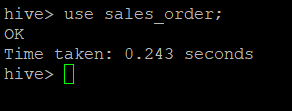


**Step 5**: Create database in hive.

create database sales\_order;



**Step 6**: Use hive\_assignment, to create the table in particular database.



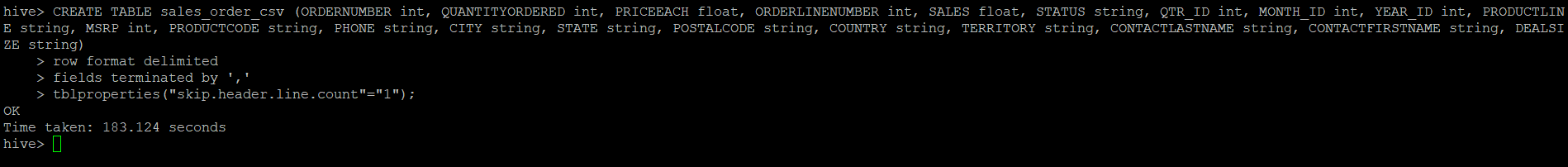
**Step 7**: Create table sales\_order\_csv.

CREATE TABLE sales\_order\_csv (ORDERNUMBER int, QUANTITYORDERED int, PRICEEACH float, ORDERLINENUMBER int, SALES float, STATUS string, QTR\_ID int, MONTH\_ID int, YEAR\_ID int, PRODUCTLINE string, MSRP int, PRODUCTCODE string, PHONE string, CITY string, STATE string, POSTALCODE string, COUNTRY string, TERRITORY string, CONTACTLASTNAME string, CONTACTFIRSTNAME string, DEALSIZE string)

row format delimited

fields terminated by ','

tblproperties("skip.header.line.count"="1"); --this is uses to skip the first line in csv file.



**Step 8:** Load Data into table.

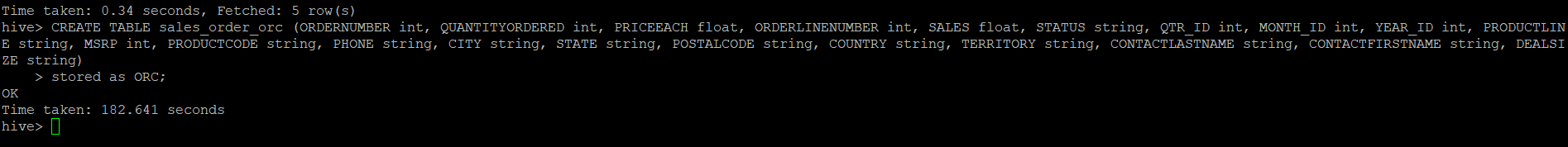
LOAD DATA INPATH '/apps/hive/data/sales\_order\_data.csv' into table sales\_order\_csv;



**Step 9**: Create ORC format table.

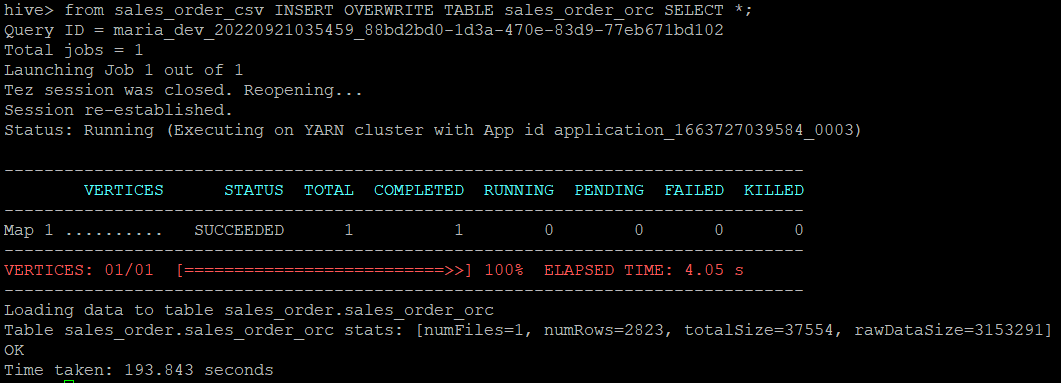
CREATE TABLE sales\_order\_orc (ORDERNUMBER int, QUANTITYORDERED int, PRICEEACH float, ORDERLINENUMBER int, SALES float, STATUS string, QTR\_ID int, MONTH\_ID int, YEAR\_ID int, PRODUCTLINE string, MSRP int, PRODUCTCODE string, PHONE string, CITY string, STATE string, POSTALCODE string, COUNTRY string, TERRITORY string, CONTACTLASTNAME string, CONTACTFIRSTNAME string, DEALSIZE string)

stored as ORC;



**Step 10**: inserting data in ORC format.

from sales\_order\_csv INSERT OVERWRITE TABLE sales\_order\_orc SELECT \*;

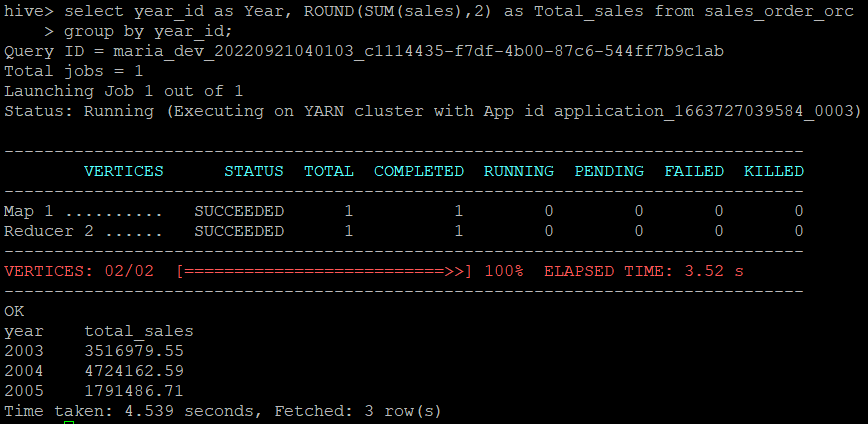
****

**Now let’s start writing queries from table “sales\_order\_orc”**

1. **Calculate total sales per year**

select year\_id as Year, ROUND(SUM(sales),2) as Total\_sales from sales\_order\_orc

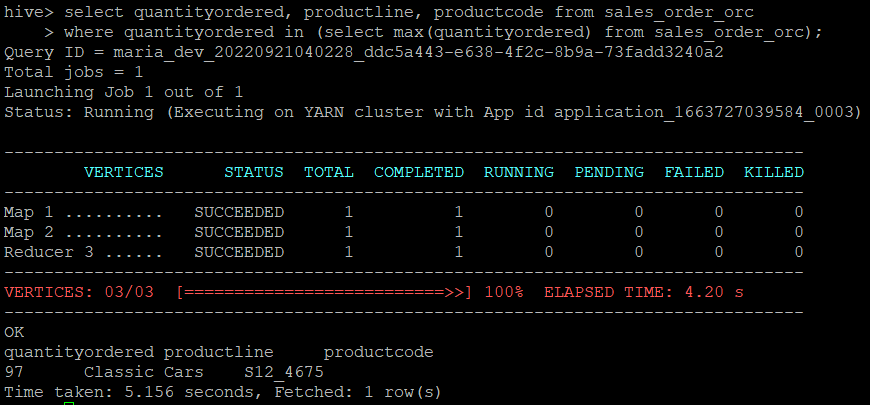
group by year\_id;



1. **Find a product for which maximum orders were placed**

select quantityordered, productline, productcode from sales\_order\_orc

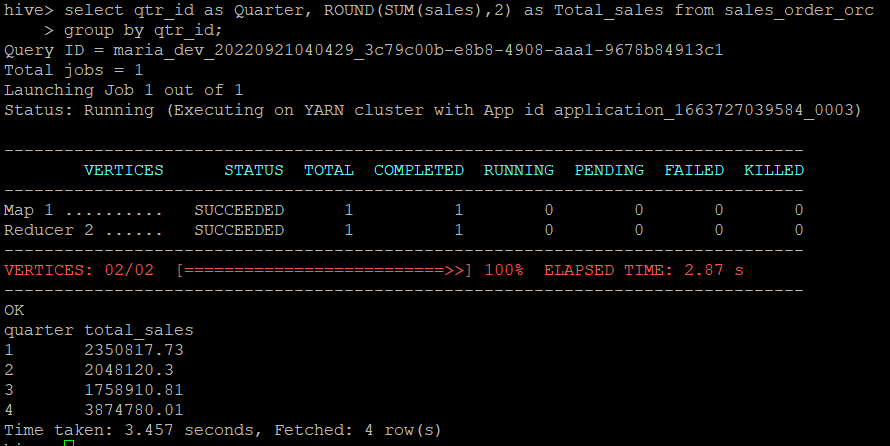
where quantityordered in (select max(quantityordered) from sales\_order\_orc);



1. **Calculate the total sales for each quarter**

select qtr\_id as Quarter, ROUND(SUM(sales),2) as Total\_sales from sales\_order\_orc

group by qtr\_id;

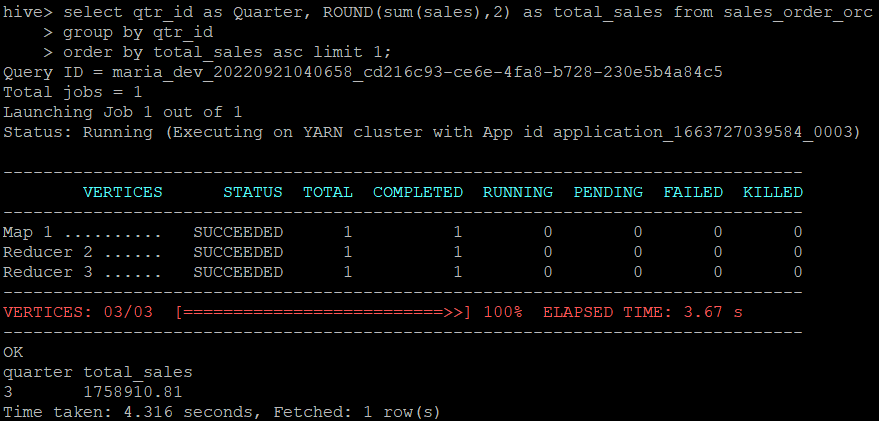


1. **In which quarter sales was minimum**

select qtr\_id as Quarter, ROUND(sum(sales),2) as total\_sales from sales\_order\_orc

group by qtr\_id

order by total\_sales asc limit 1;



1. **In which country sales was maximum and in which country sales was minimum**

select \*

from(

select country, ROUND(sum(sales),2) as min\_sales

from sales\_order\_orc

group by country

order by min\_sales asc limit 1) tmp

union

select \*

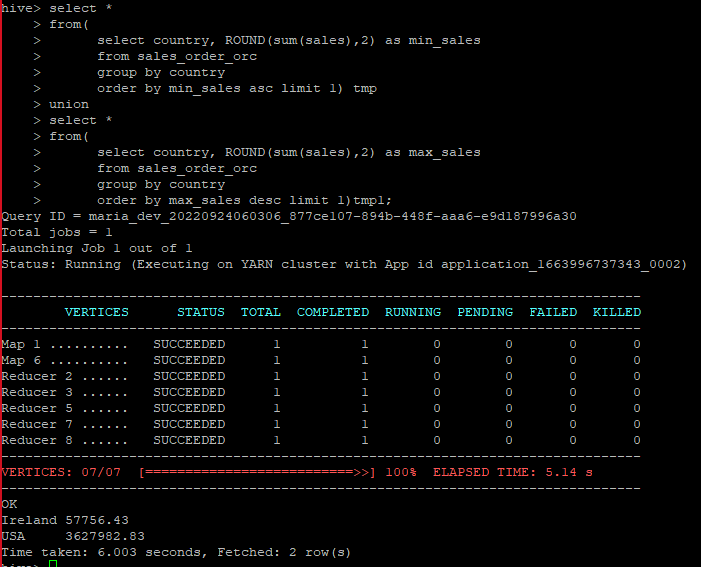
from(

select country, ROUND(sum(sales),2) as max\_sales

from sales\_order\_orc

group by country

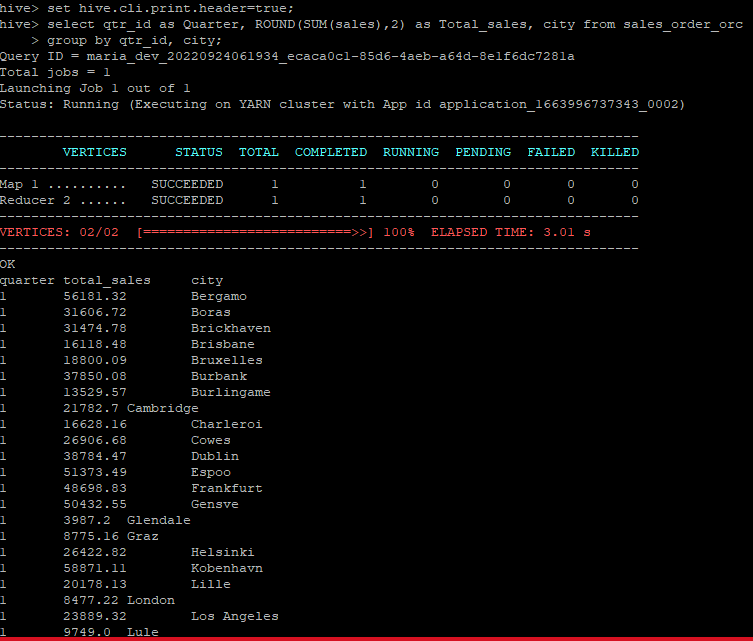
order by max\_sales desc limit 1) tmp1;



1. **Calculate quarterly sales for each city**

select qtr\_id as Quarter, ROUND(SUM(sales),2) as Total\_sales, city from sales\_order\_orc

group by qtr\_id, city;



1. **Find a month for each year in which maximum number of quantities were sold**

select year\_id, month\_id, quantityordered from sales\_order\_orc a

where quantityordered in (select max(quantityordered) from sales\_order\_orc b where a.year\_id=b.year\_id )

group by year\_id, month\_id, quantityordered;

