**HIVE ASSIGNMENT-1**

**Question Statement:**

1. Download vehicle sales data -> https://github.com/shashank-mishra219/Hive-Class/blob/main/sales\_order\_data.csv

2. Store raw data into hdfs location

3. Create a internal hive table "sales\_order\_csv" which will store csv data sales\_order\_csv .. make sure to skip header row while creating table

4. Load data from hdfs path into "sales\_order\_csv"

5. Create an internal hive table which will store data in ORC format "sales\_order\_orc"

6. Load data from "sales\_order\_csv" into "sales\_order\_orc"

Perform below menioned queries on "sales\_order\_orc" table :

a. Calculatye total sales per year

b. Find a product for which maximum orders were placed

c. Calculate the total sales for each quarter

d. In which quarter sales was minimum

e. In which country sales was maximum and in which country sales was minimum

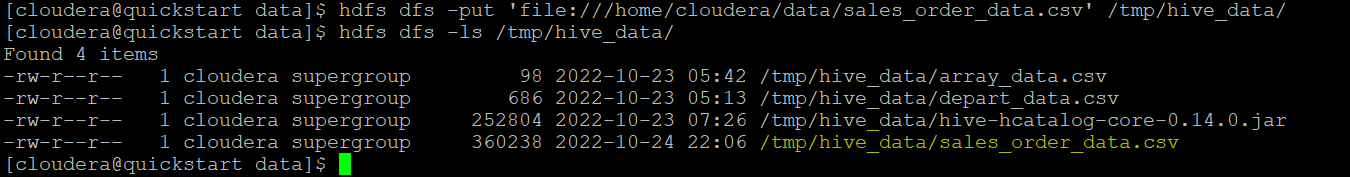
f. Calculate quartelry sales for each city

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

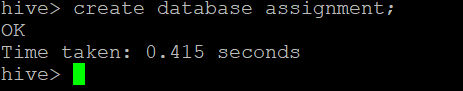
Data Preparation:

Copied the file to HDFS location:

**hdfs dfs -put 'file:///home/cloudera/data/sales\_order\_data.csv' /tmp/hive\_data/**



Create separate database for assignments.



Create table sales\_data;

DDL :

create table sales\_order\_data(

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,

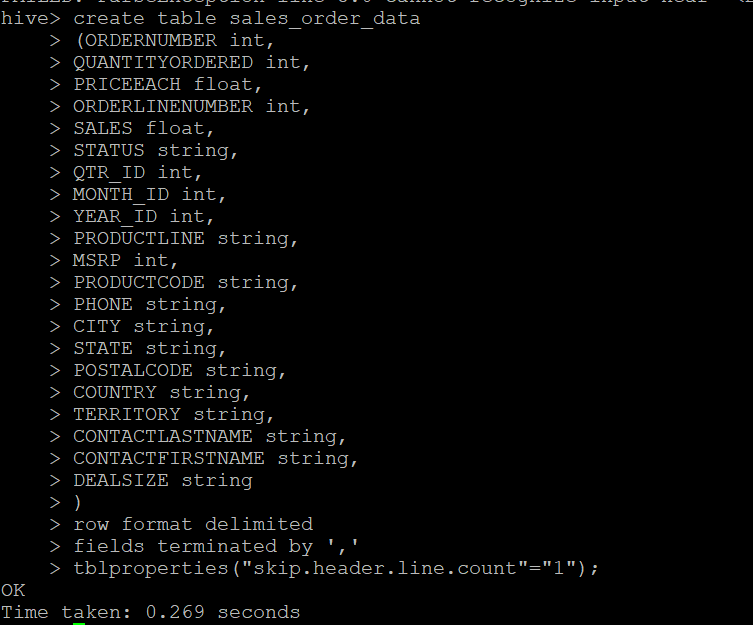
DEALSIZE string

)

row format delimited

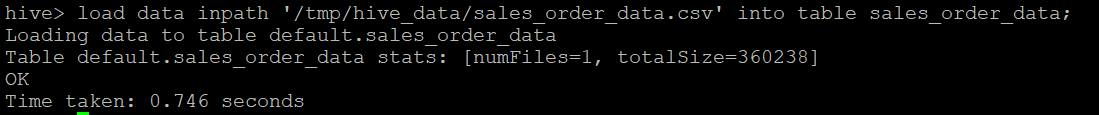
fields terminated by ','

tblproperties("skip.header.line.count"="1");



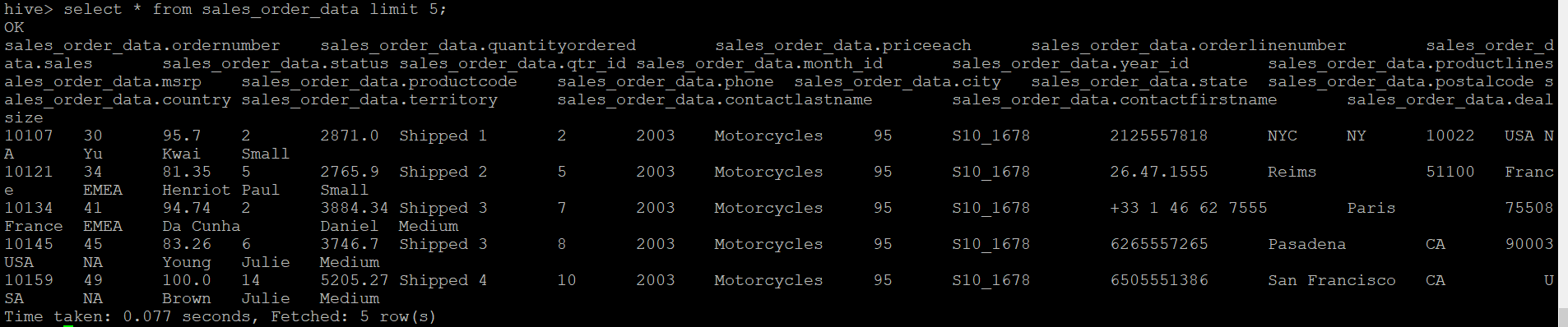
Load data from HDFS location into table sales\_order\_data.

**load data inpath '/tmp/hive\_data/sales\_order\_data.csv' into table sales\_order\_data;**



set hive.cli.print.header=true;

select \* from sales\_order\_data limit 5;



Create a ORC Table:

create table sales\_order\_data\_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,

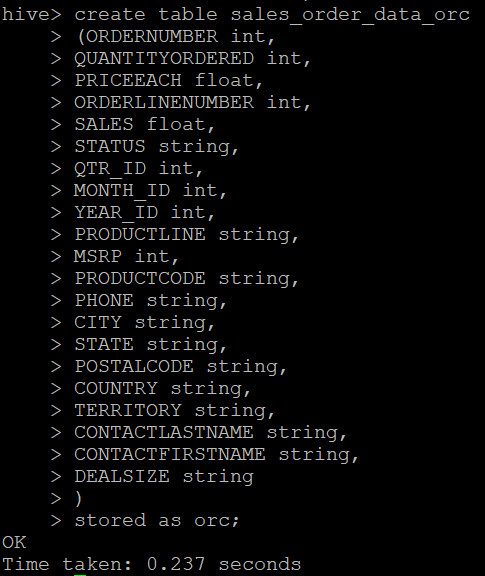
CONTACTLASTNAME string,

CONTACTFIRSTNAME string,

DEALSIZE string

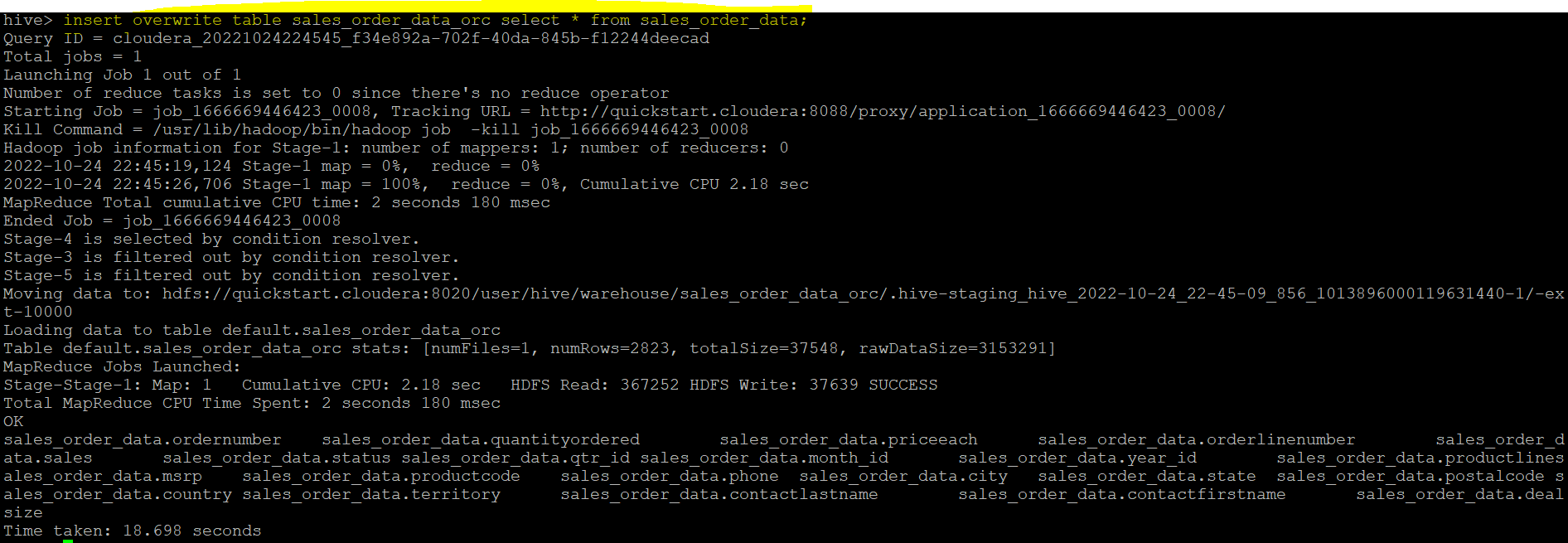
)

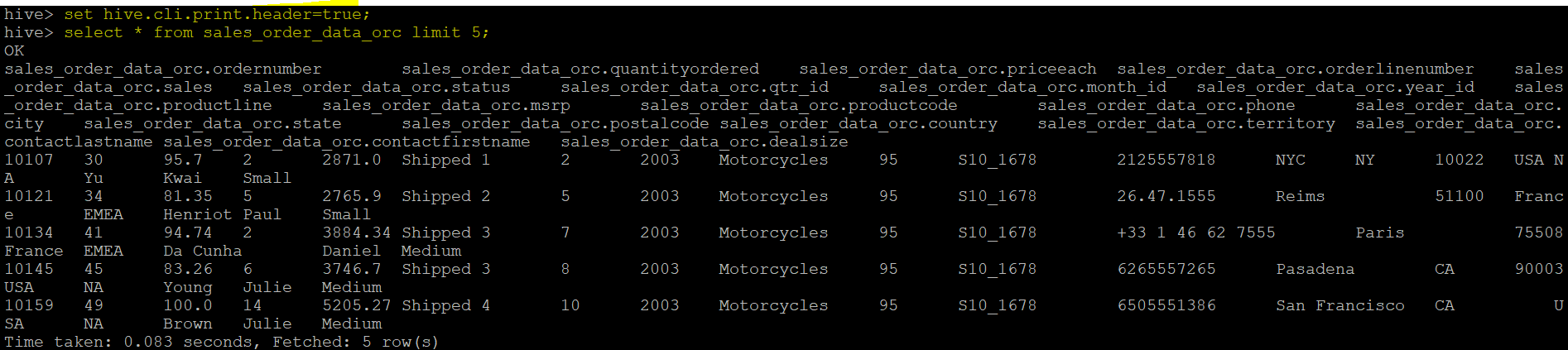
Stored as orc;



Load data from **sales\_order\_data** into **sales\_order\_data\_orc**

**insert overwrite table sales\_order\_data\_orc select \* from sales\_order\_data;**

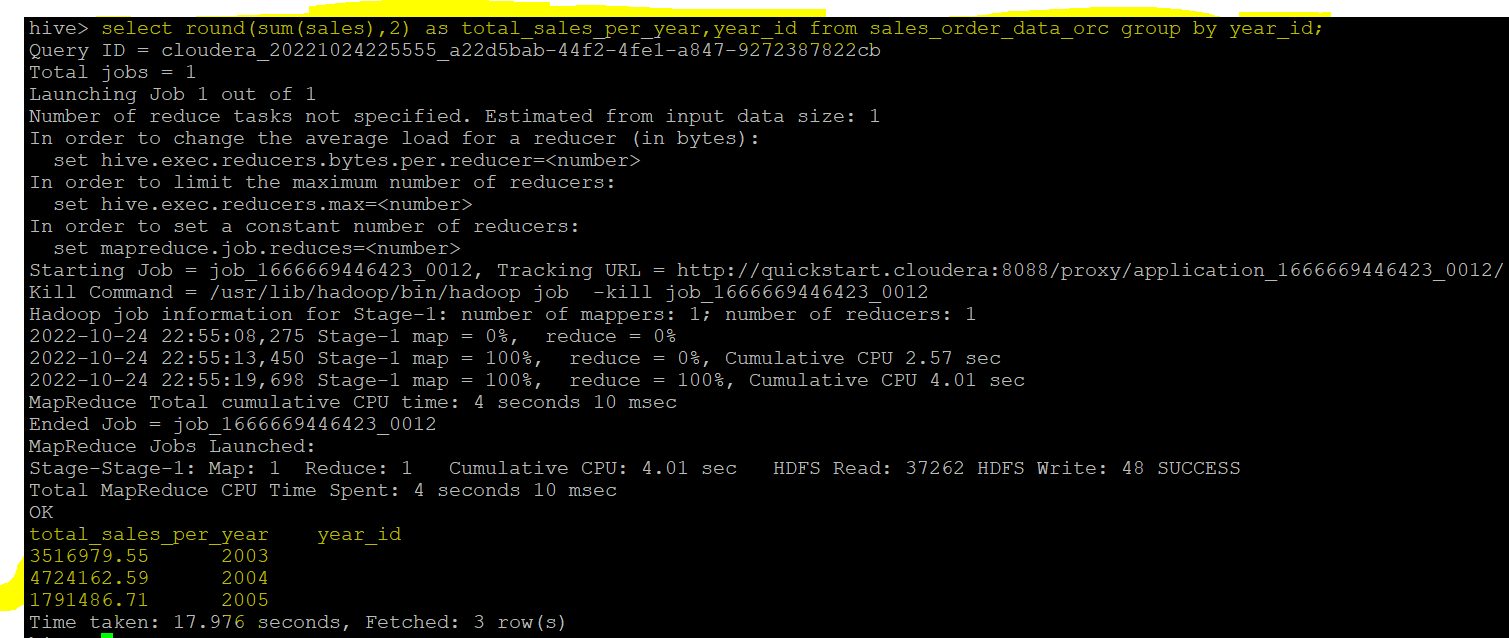




Perform below mentioned queries on "sales\_order\_data\_orc" table

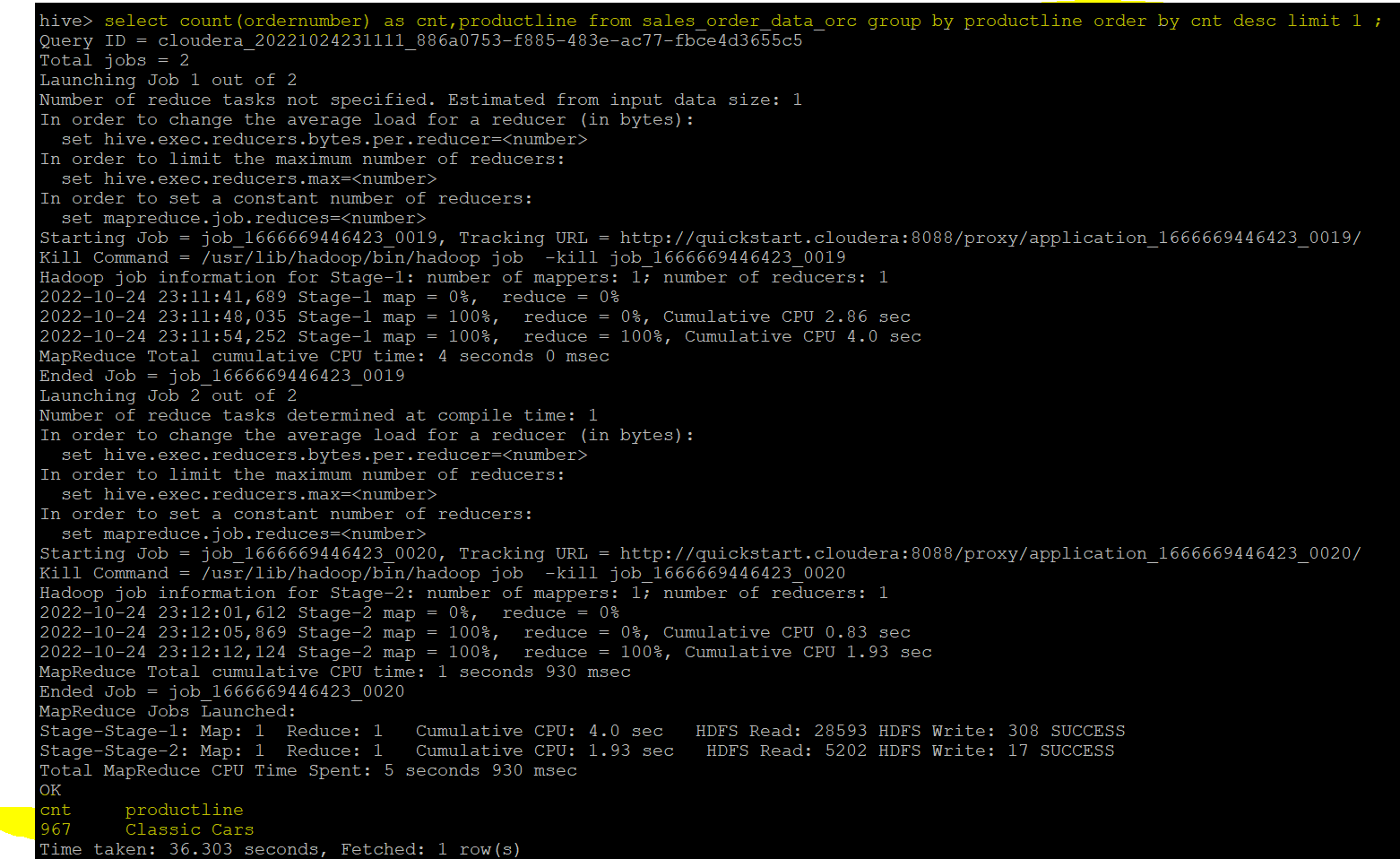
1. **Calculate total sales per year**

**select round(sum(sales),2) as total\_sales\_per\_year,year\_id from sales\_order\_data\_orc group by year\_id;**



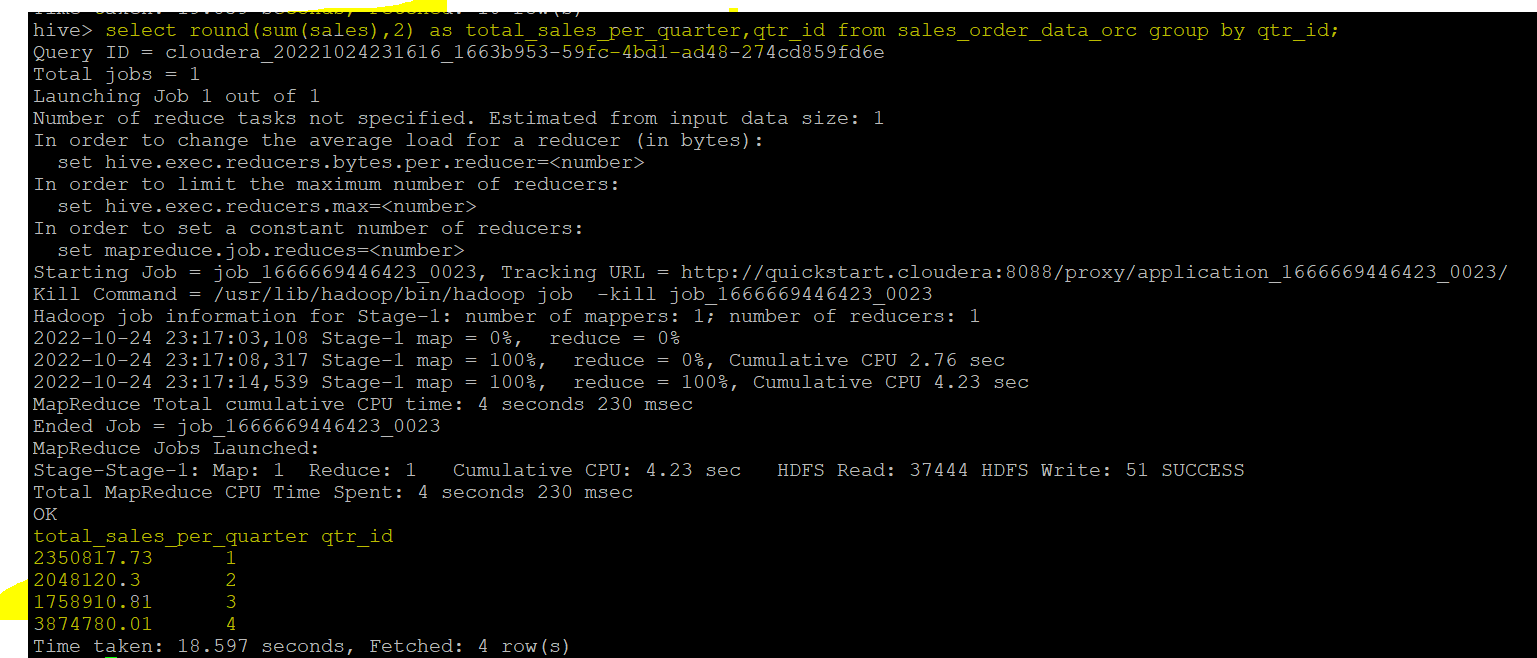
1. **Find a product for which maximum orders were placed**

**select count(ordernumber) as cnt,productline from sales\_order\_data\_orc group by productline order by cnt desc limit 1 ;**



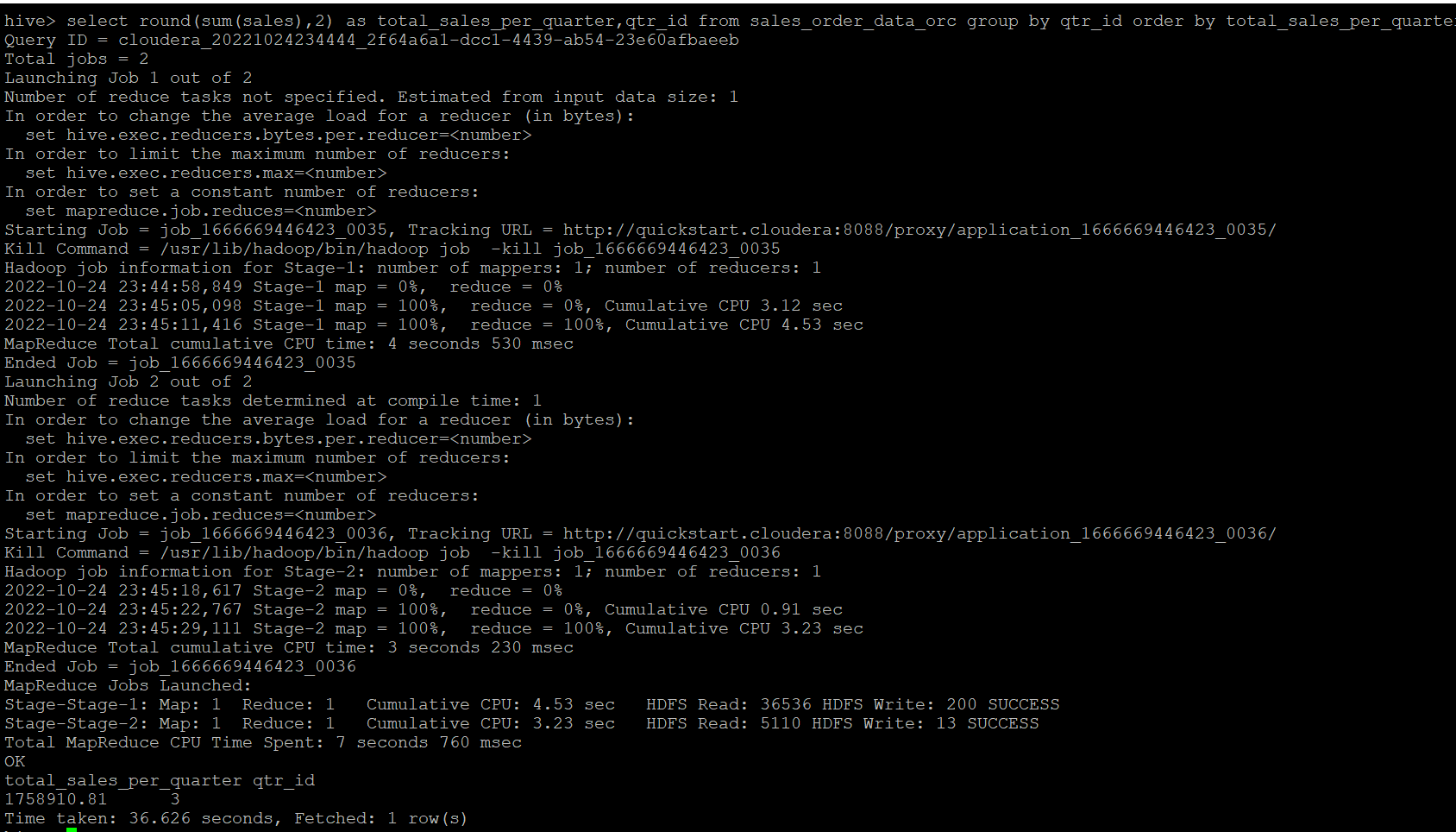
1. **Calculate the total sales for each quarter**

**select round(sum(sales),2) as total\_sales\_per\_quarter,qtr\_id from sales\_order\_data\_orc group by qtr\_id;**



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

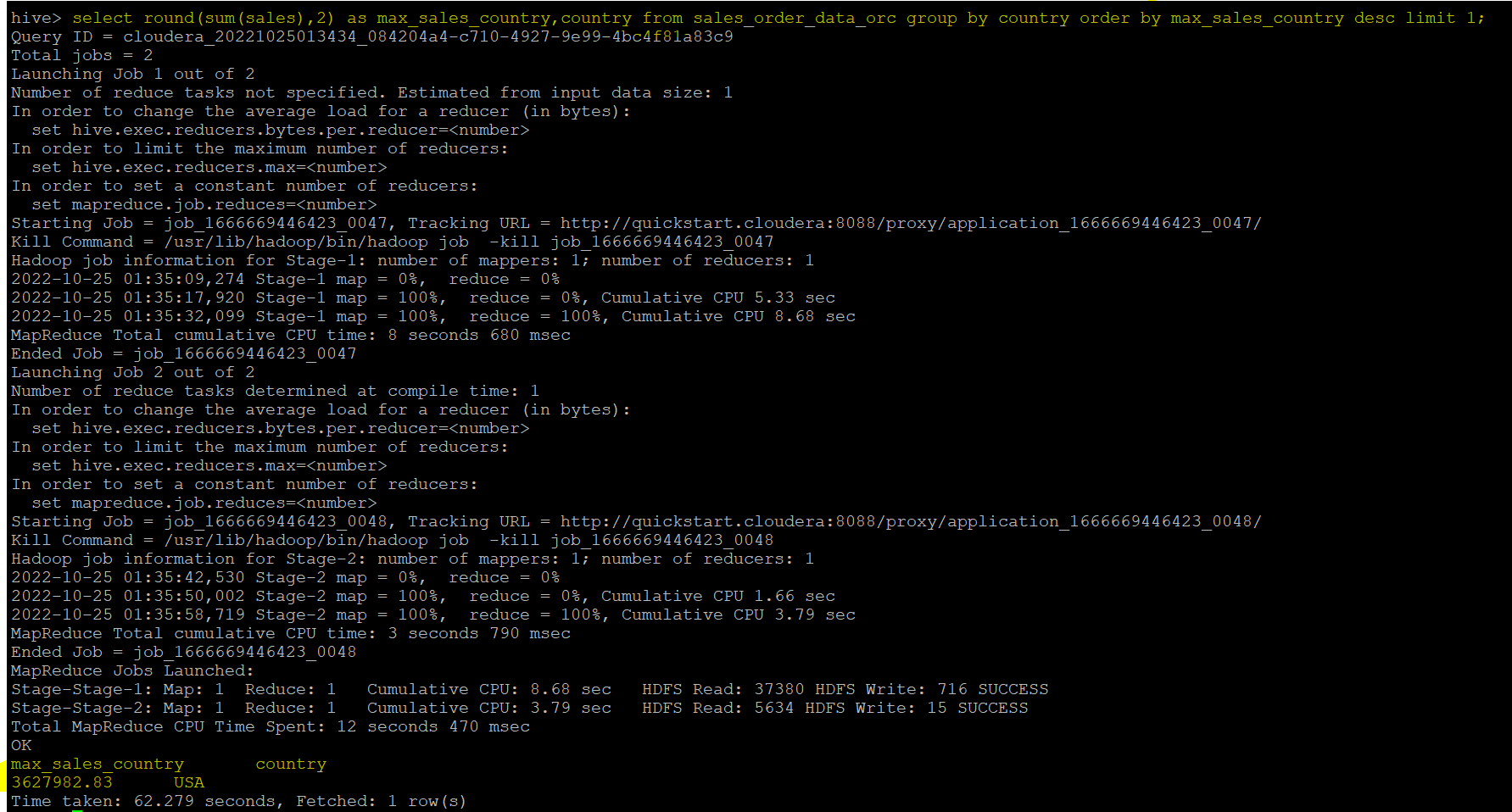
**select round(sum(sales),2) as total\_sales\_per\_quarter,qtr\_id from sales\_order\_data\_orc group by qtr\_id order by total\_sales\_per\_quarter limit 1;**



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

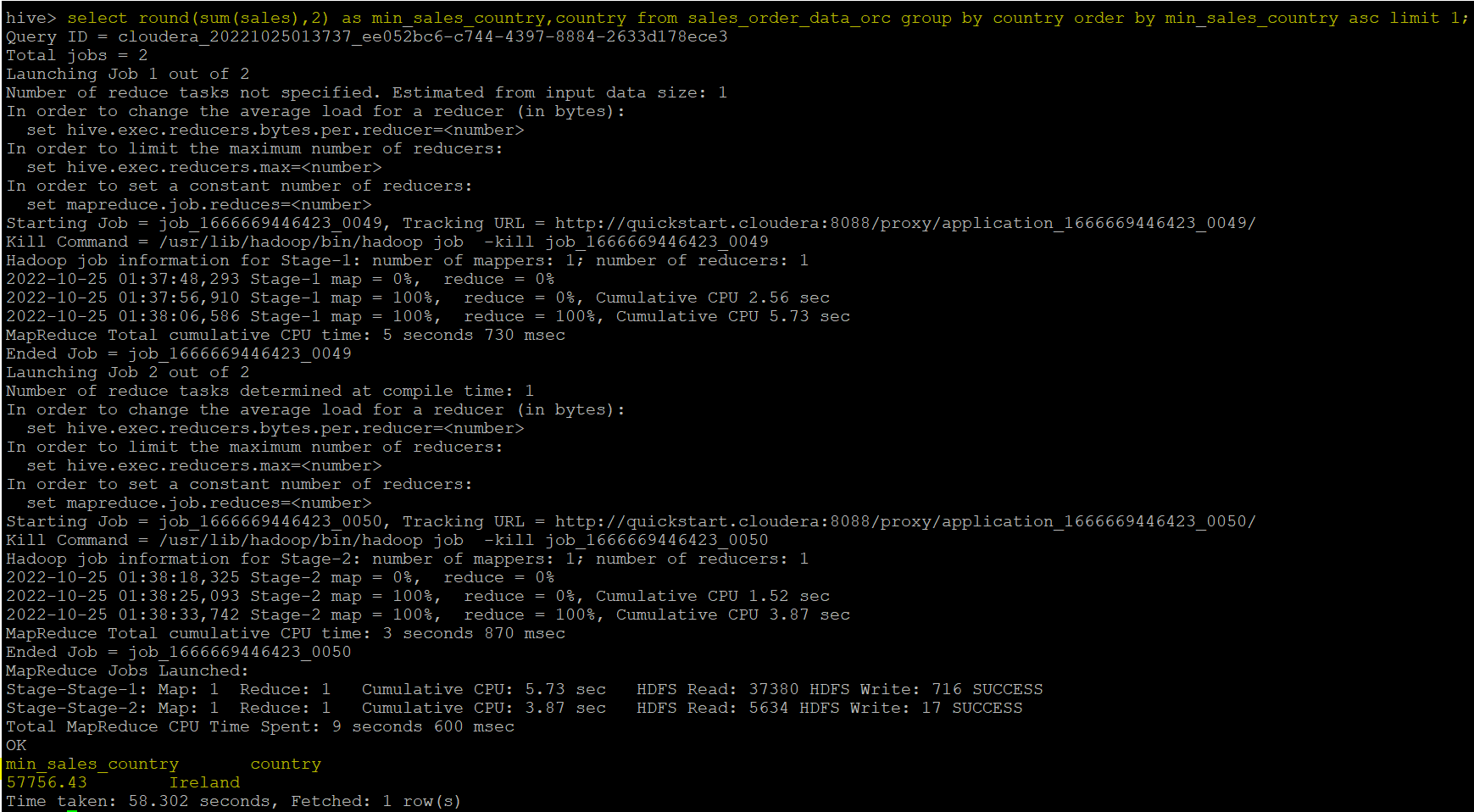
**Maximum:**

**select round(sum(sales),2) as max\_sales\_country,country from sales\_order\_data\_orc group by country order by max\_sales\_country desc limit 1;**



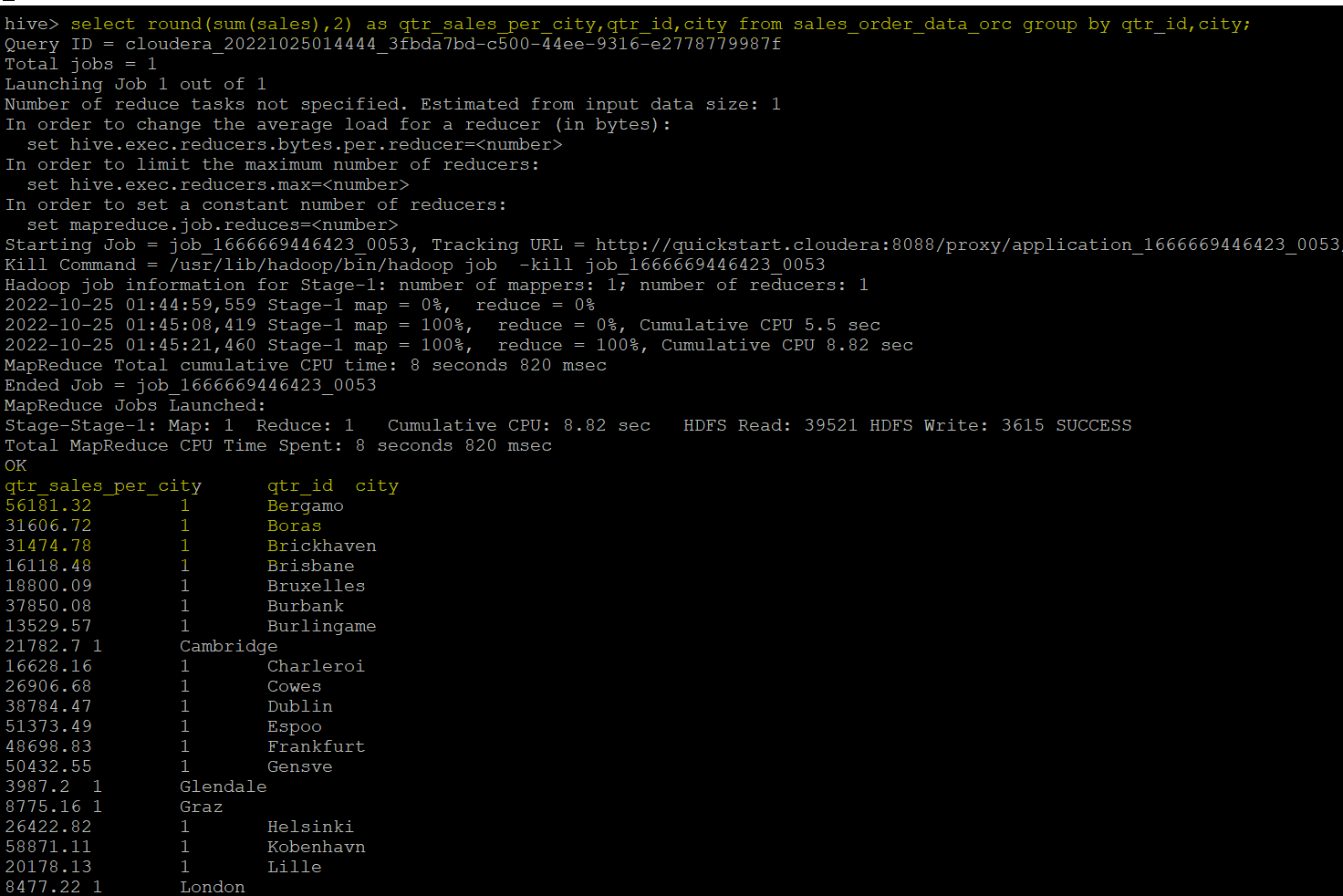
**Minimum:**

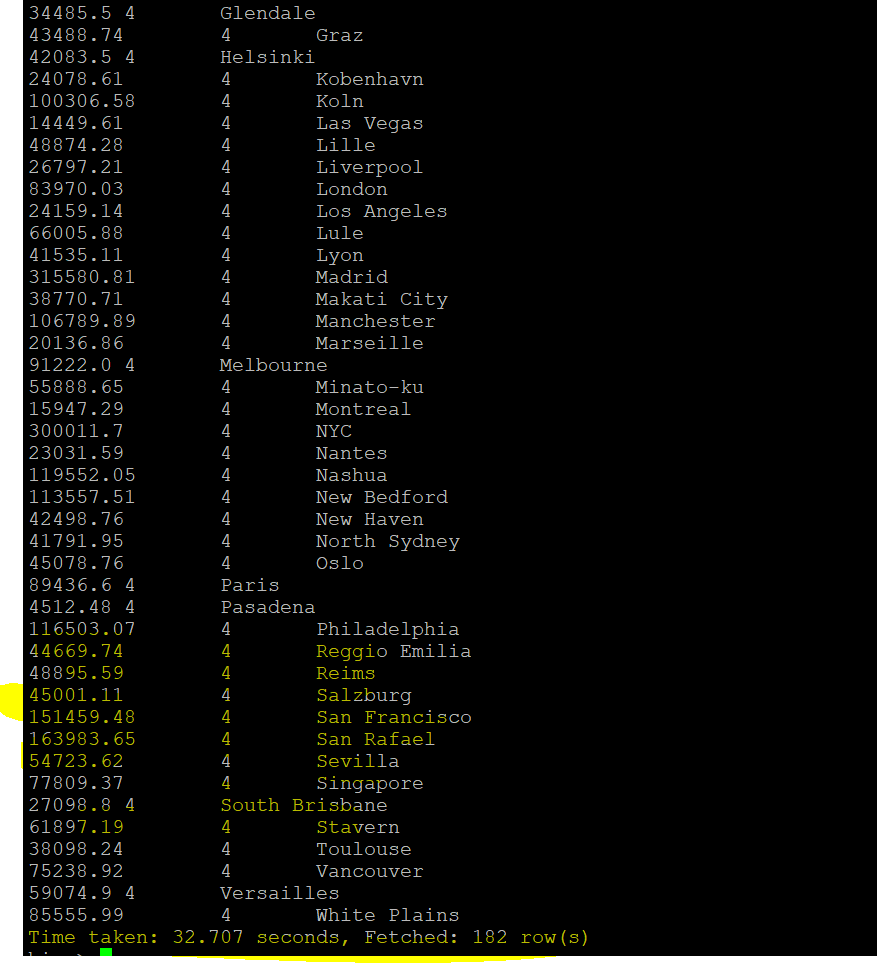
**select round(sum(sales),2) as min\_sales\_country,country from sales\_order\_data\_orc group by country order by min\_sales\_country asc limit 1;**



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

**select round(sum(sales),2) as qtr\_sales\_per\_city,qtr\_id,city from sales\_order\_data\_orc group by qtr\_id,city;**





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

**select b.max\_qty\_per\_month\_year as max\_qty\_per\_month\_year,b.month\_id,b.year\_id from**

**(select a.\*, rank () over (partition by a.year\_id order by a.max\_qty\_per\_month\_year desc) as rnk from**

**(select sum(quantityordered) as max\_qty\_per\_month\_year,month\_id,year\_id from sales\_order\_data\_orc group by month\_id,year\_id order by max\_qty\_per\_month\_year desc) a)b**

**where b.rnk=1**

