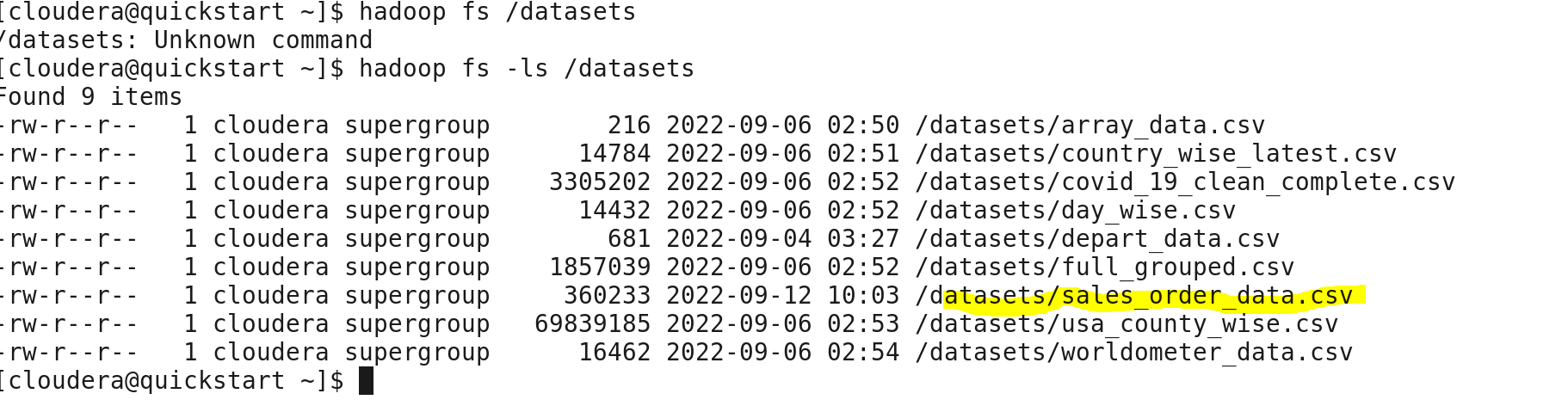
Hive Solution

Q2) Store raw data into hdfs location

Ans)

* Hadoop fs –put /home/cloudera/dataset/sales\_order\_data.csv /datasets
* Hadoop fs – ls /datasets

Output -:

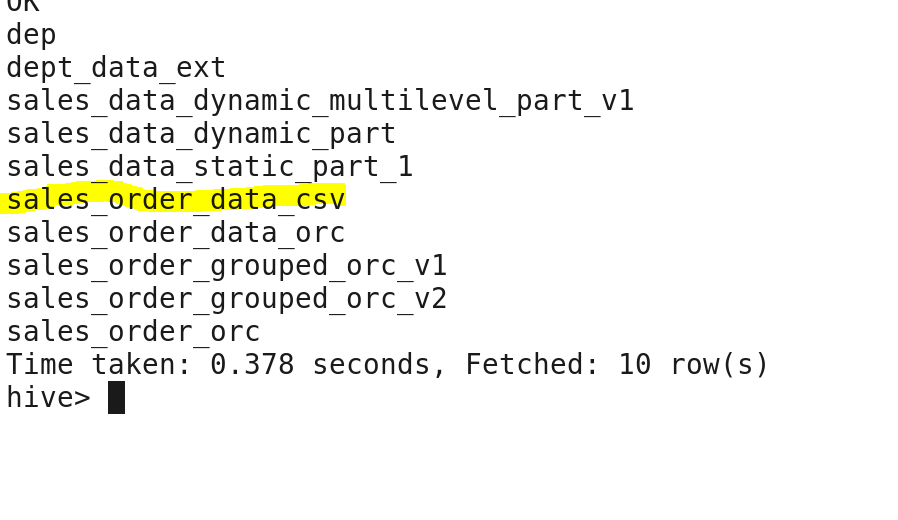


Q3) Create a internal hive table "sales\_order\_csv" which will store csv data sales\_order\_csv .. make sure to skip header row while creating table

ANS)

* Use hive\_class\_r1;
* create table sales\_order\_data\_csv
* (
* order\_number int,
* quantity\_ordered int,
* price\_each float,
* order\_line\_number int,
* sales float,
* status string,
* qtr\_id int,
* month\_id int,
* year\_id int,
* product\_line string,
* msrp int,
* product\_line string,
* phone string,
* city string,
* state string,
* postal\_code string,
* country string,
* territory string,
* contact\_last\_name string,
* contact\_first\_name string,
* deal\_size string
* )
* row format delimited
* fields terminated by ','
* tblproperties("skip.header.line.count"="1")
* ;

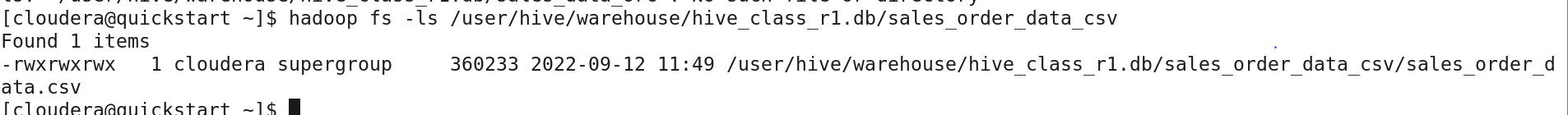
Output - :



Q4) Load data from hdfs path into "sales\_order\_csv"

Load data inpath “/datasets/sales\_order\_data.csv” into table sales\_order\_data\_csv

Output - :

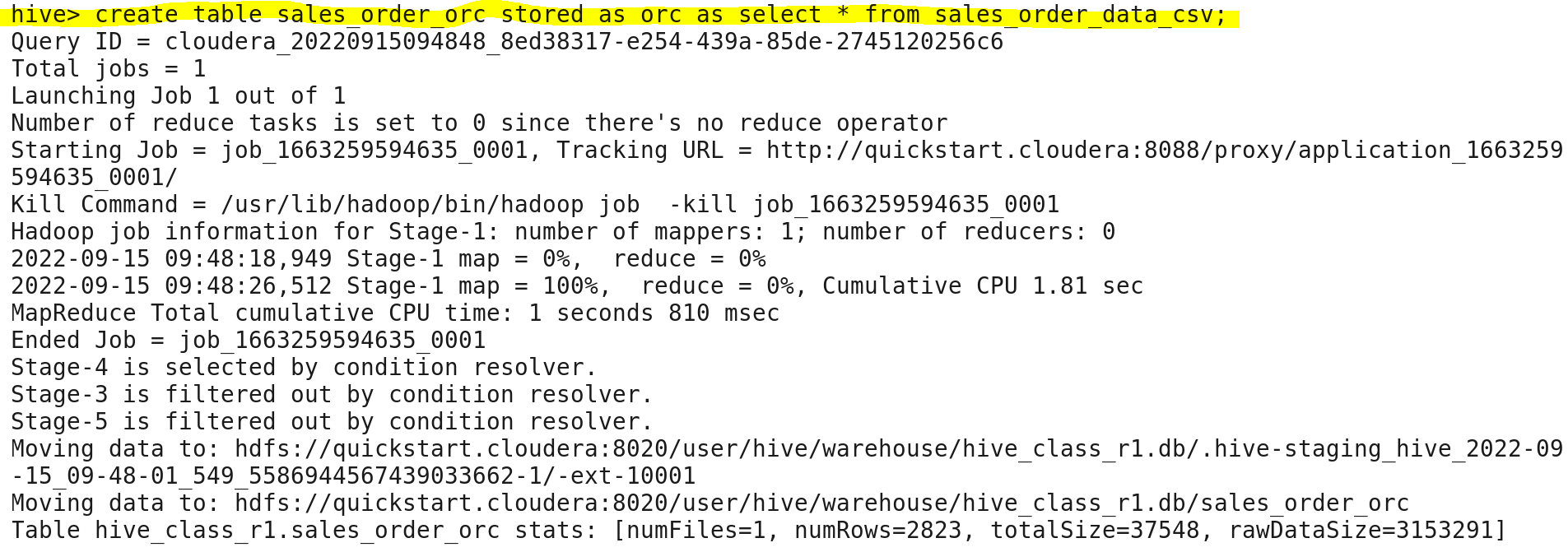


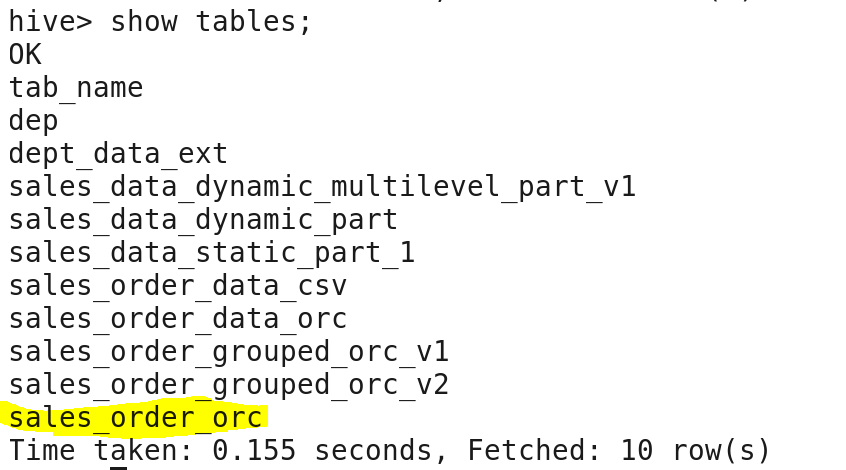
Q5) Load data from "sales\_order\_csv" into "sales\_order\_orc"

ANS)

* create table sales\_order\_orc stored as orc as select \* from sales\_order\_data\_csv;

output - :





Q6) Load data from "sales\_order\_csv" into "sales\_order\_orc"

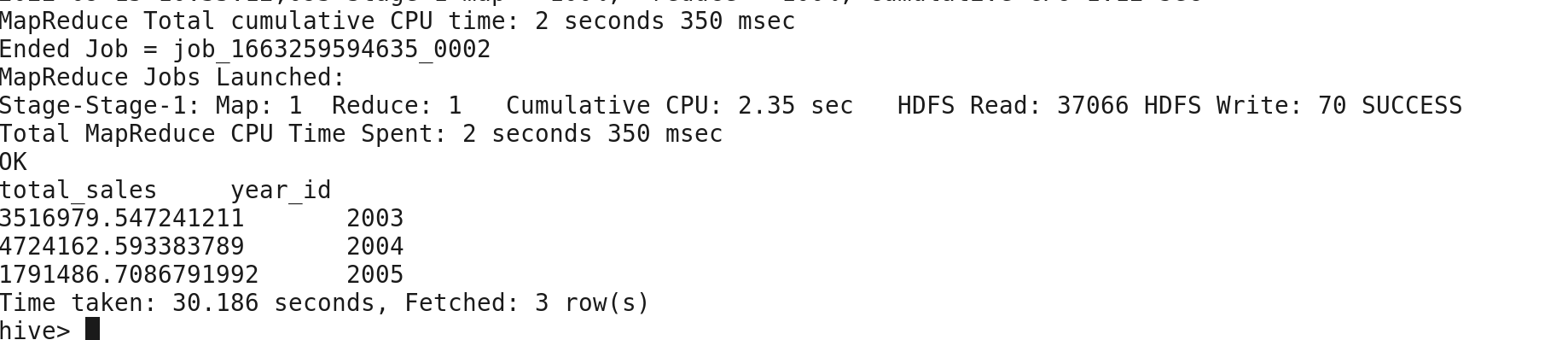
* from sales\_order\_data\_csv insert overwrite table sales\_order\_data\_orc;
* Perform below menioned queries on "sales\_order\_orc" table :

1. Calculate total sales per year

Ans)

* Select sum(sales) as total\_sales ,year\_id from sales\_order\_orc
* Group by year\_id;

Output - :

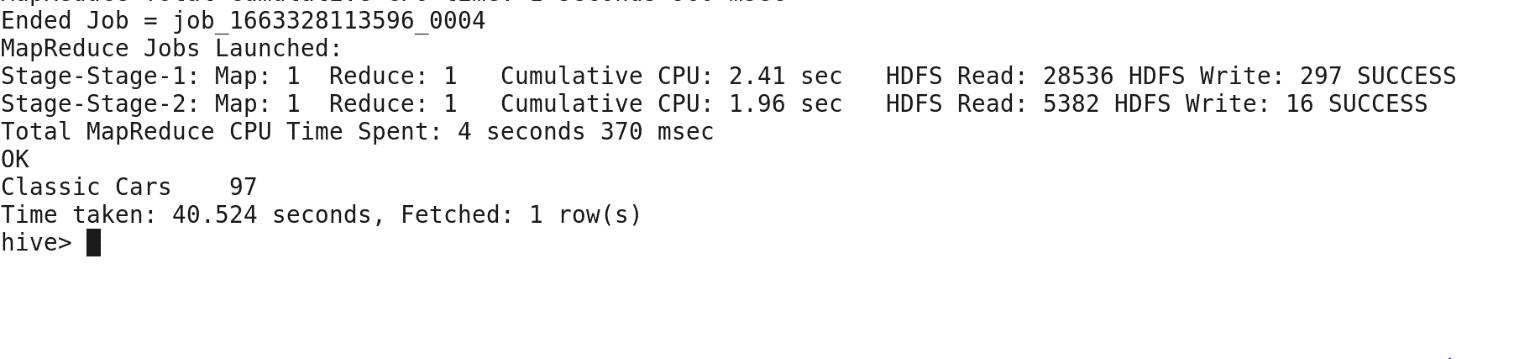


1. Find a product for which maximum orders were placed

Ans)

* select product\_line,max(quantity\_ordered) as max\_order from sales\_order\_data\_orc
* group by product\_line
* order by max\_order desc
* limit 1;

Output - :

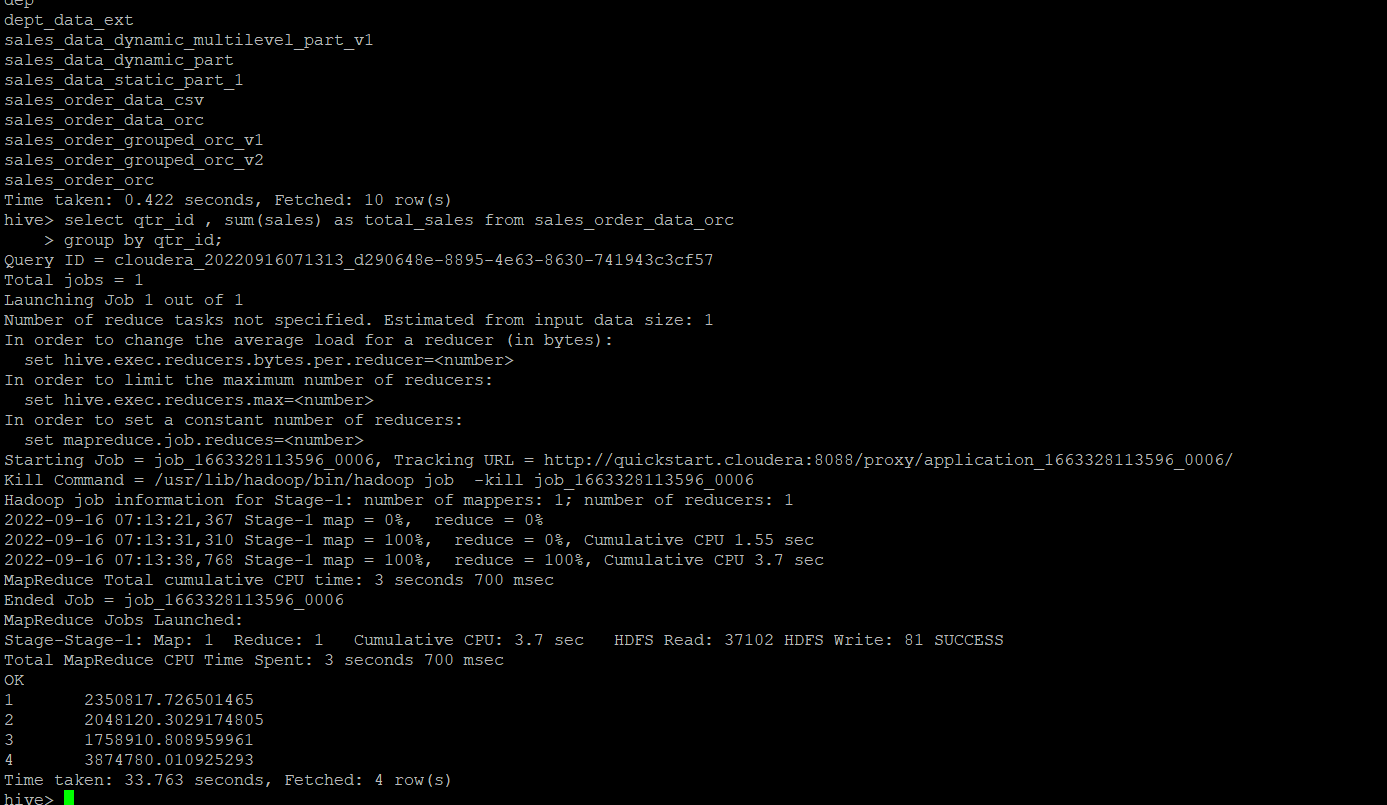


1. Calculate the total sales for each quarter

Ans)

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

Output - :



1. In which quarter sales was minimum

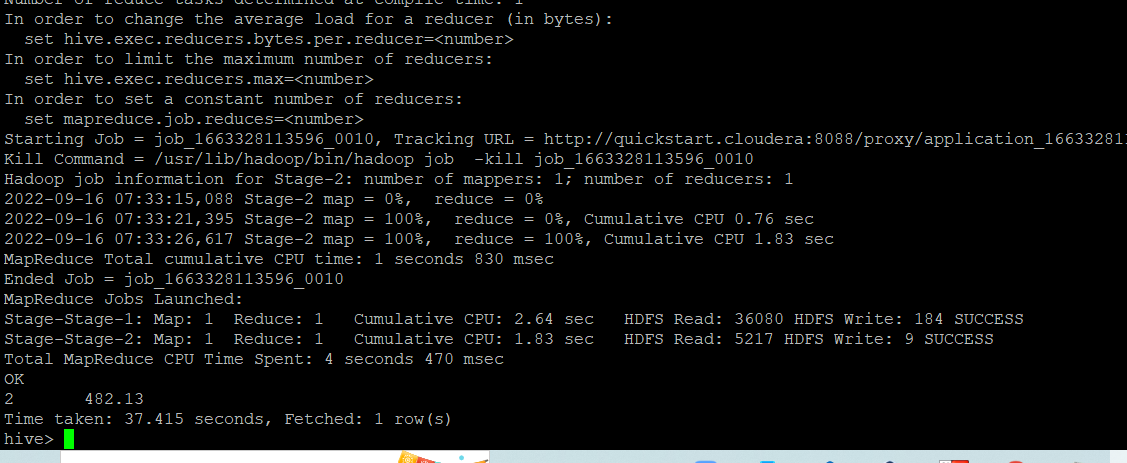
Ans ) select qtr\_id, min(sales) as min\_sales from sales\_order\_data\_csv

Group by qtr\_id

Order by min\_sales

Limit 1;

Output-:



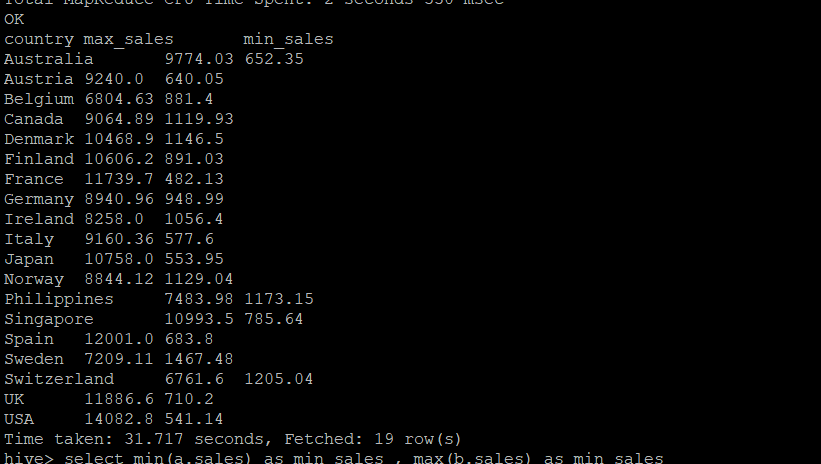
1. In which country sales was maximum and in which country sales was minimum

Ans )

select country, min(sales) as max\_sales,max(sales) as max\_sales from sales\_order\_data\_orc

group by country;

Output - :

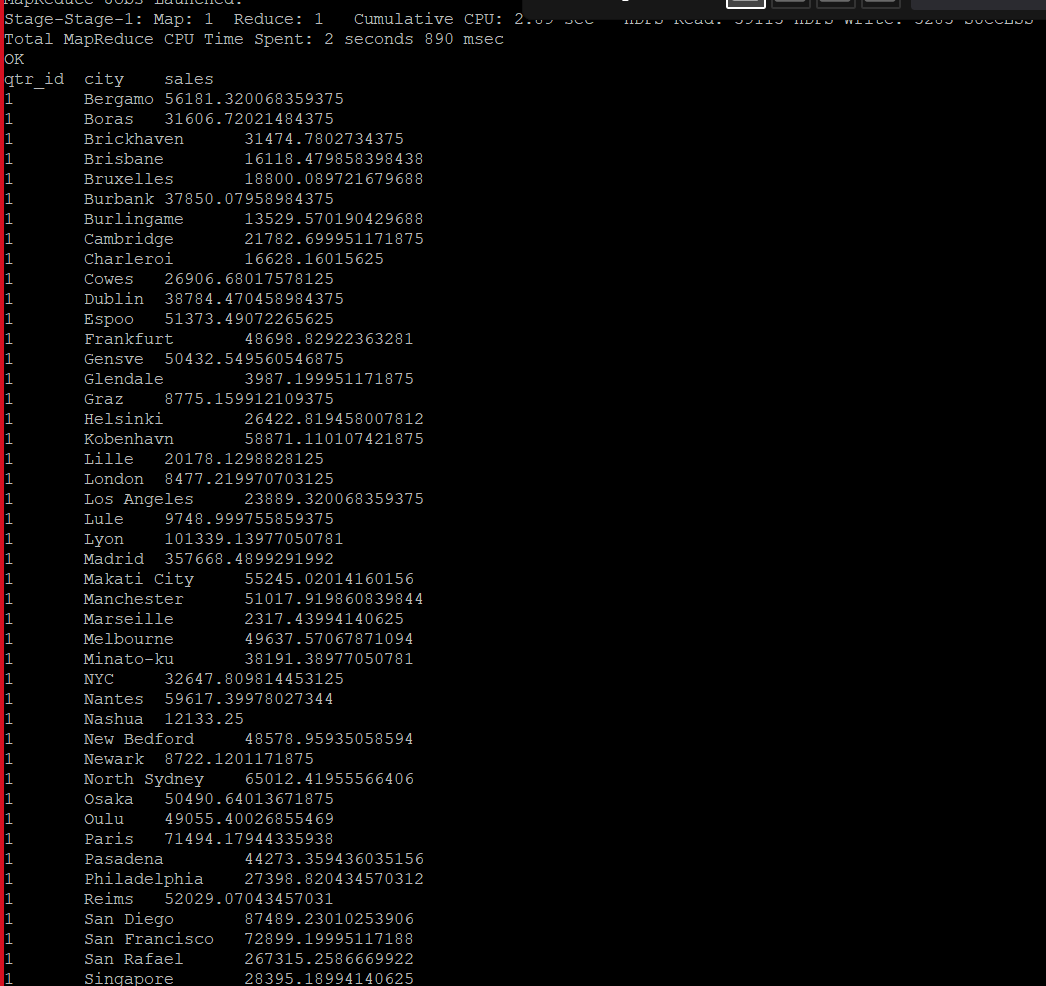


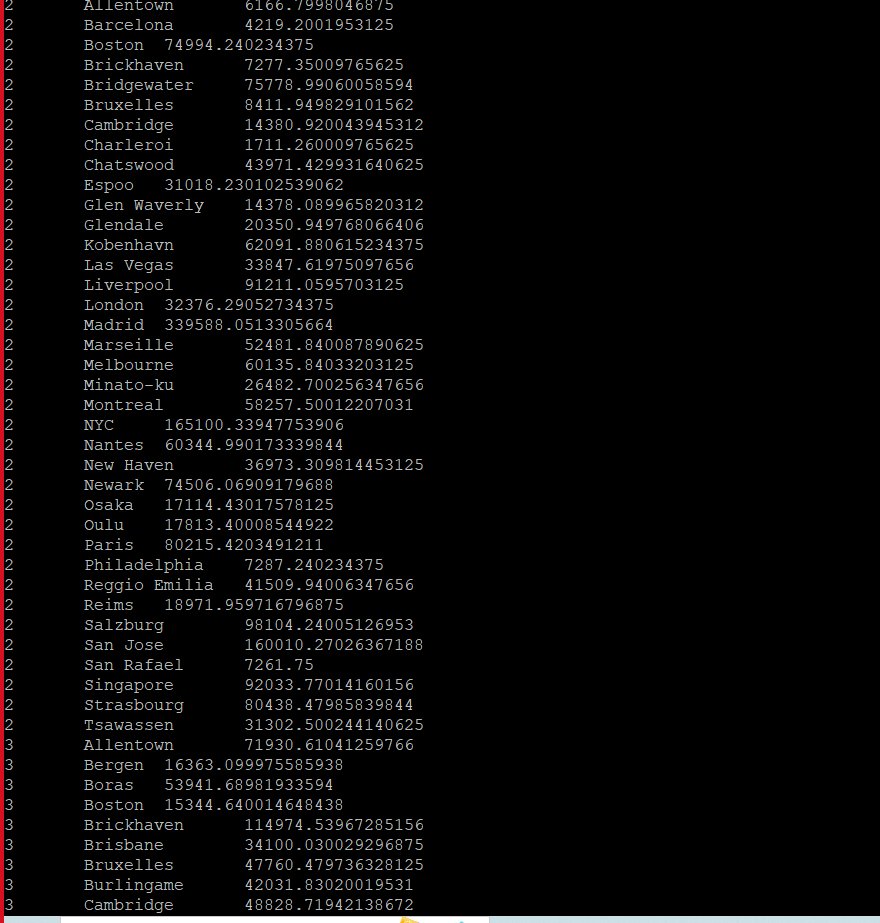
1. Calculate quarterly sales for each city

Ans)

* Select qtr\_id,country,sum(sales) as sales from sales\_order\_data\_orc
* Group by qtr\_id , country

Output - :





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

Ans)

* select year\_id,month\_id,max(quantity\_ordered) as quantity\_sold from sales\_order\_data\_orc group by year\_id,month\_id;

Output - :

