**ASSIGNMENT-1**

1. **Download vechile sales data ->** [**https://github.com/shashank-mishra219/Hive-Class/blob/main/sales\_order\_data.csv**](https://github.com/shashank-mishra219/Hive-Class/blob/main/sales_order_data.csv)

Store raw data into hdfs location

1. **Copy data from local to HDFS in directory a\_1**

hadoop fs -copyFromLocal sales\_oder\_data.csv /tmp/a\_1

1. **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**

create table sales\_order\_csv

(

ordernumber int,

orderquantity int,

priceeach float,

orderlinenumber int,

sales double,

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");

1. **Load the data from HDFS location into the sales\_order\_csv**  **table**

load data inpath '/tmp/a\_1/sales\_order\_data.csv' into table sales\_order\_csv;

1. **Create a ORC table**

create table sales\_order\_orc

(

ordernumber int,

orderquantity int,

priceeach float,

orderlinenumber int,

sales double,

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 ','

stored as orc

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

1. **load the data from sales\_order\_csv to sales\_order\_orc**

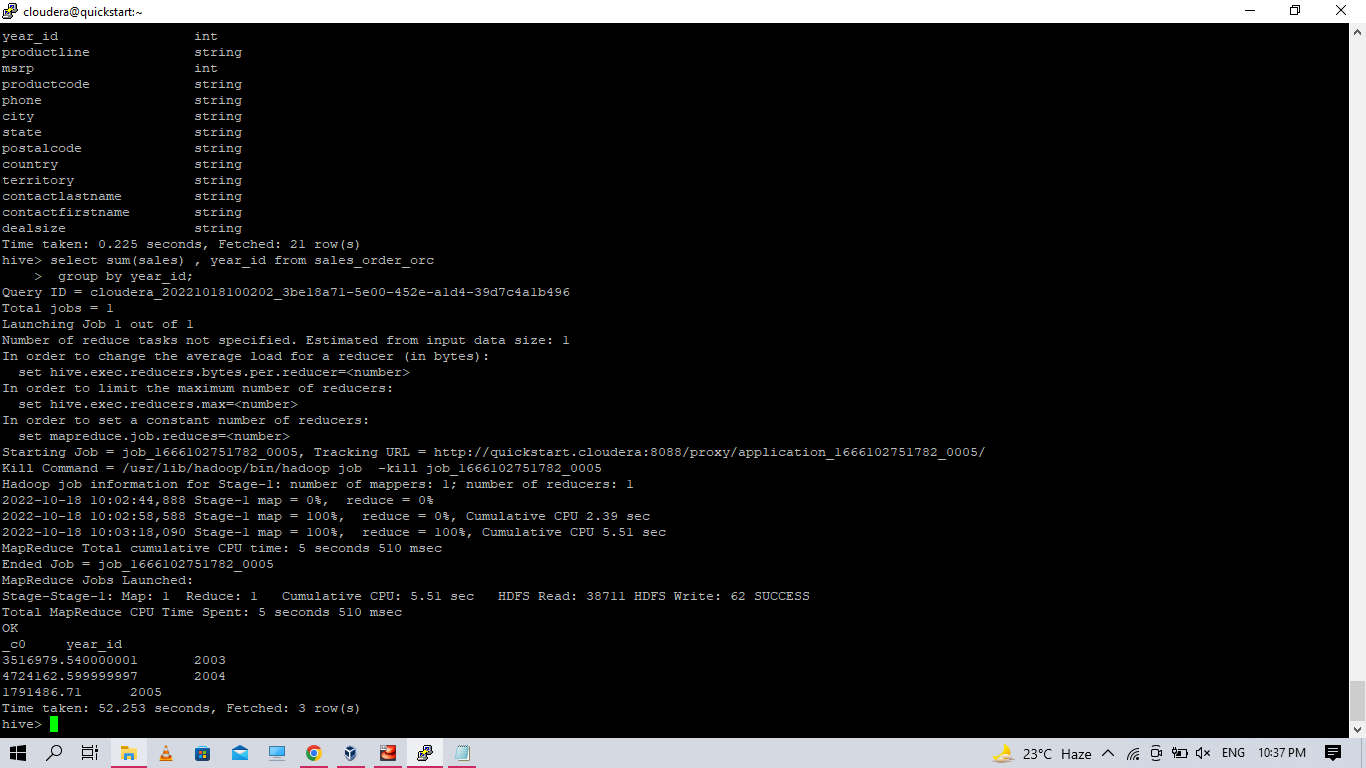
insert into sales\_order\_orc select \* from sales\_order\_csv;

**Perform the queries on "sales\_order\_orc" table :**

1. **Calculate total sales per year**

select sum(sales) , year\_id from sales\_order\_orc

group by year\_id;



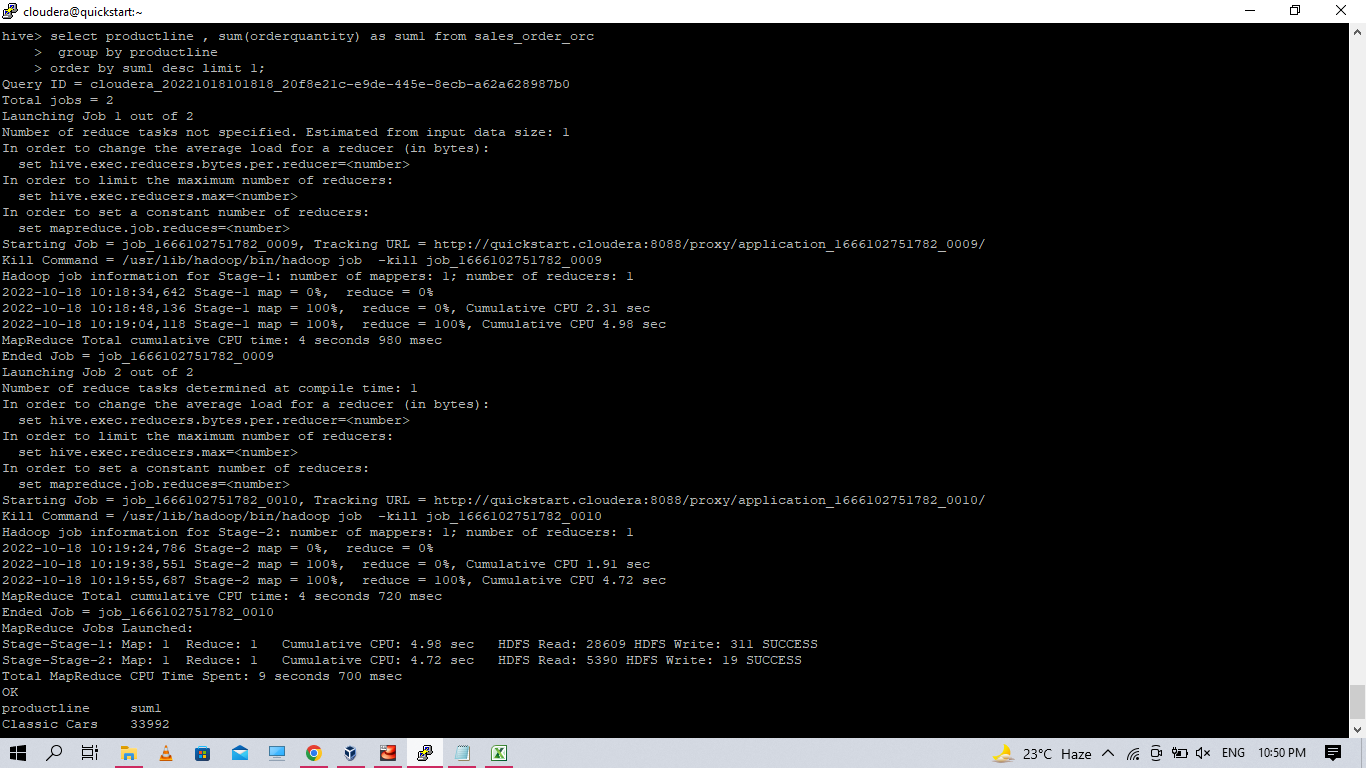
1. **Find a product for which maximum orders were placed**

select productline , sum(orderquantity) as sum1 from sales\_order\_orc

group by productline

order by sum1 desc

limit 1;

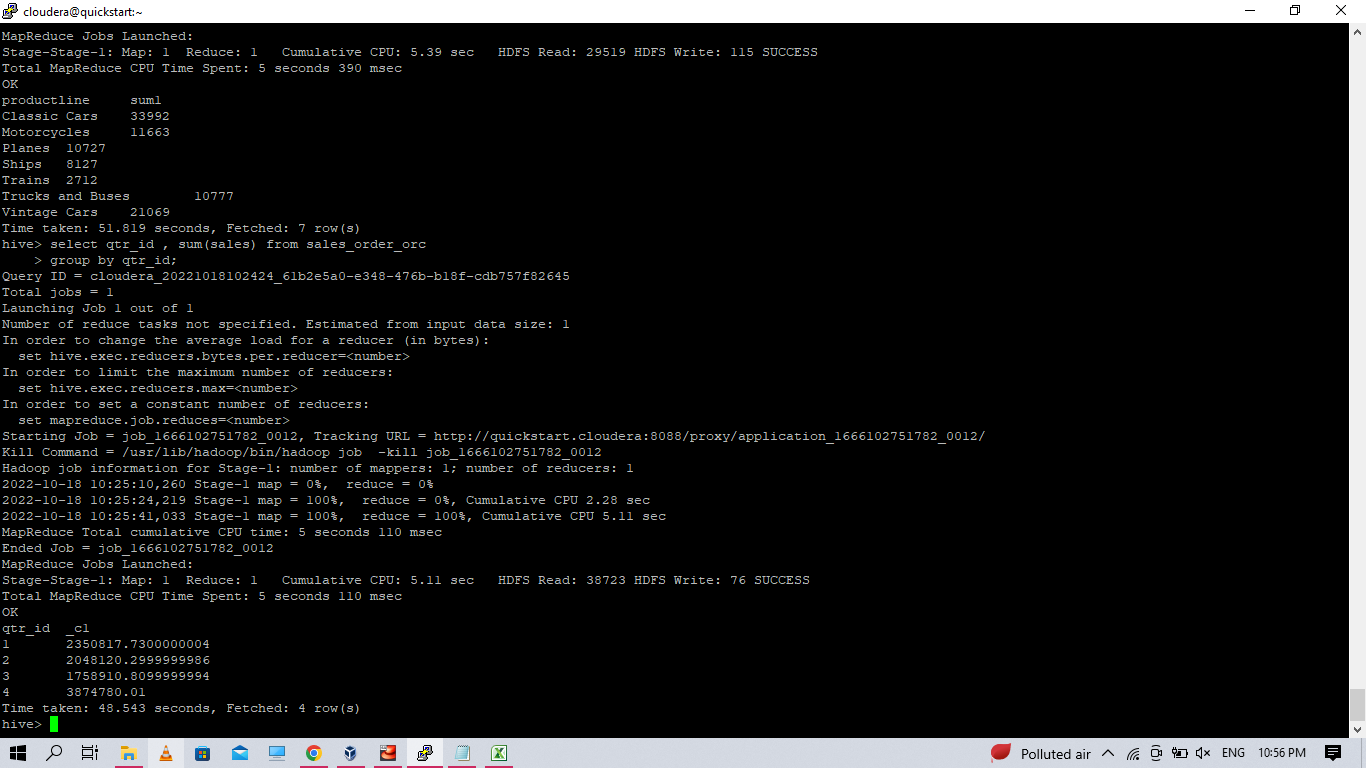


1. **Calculate the total sales for each quarter**

select sum(sales) , qtr\_id

from sales\_order\_orc

group by qtr\_id;

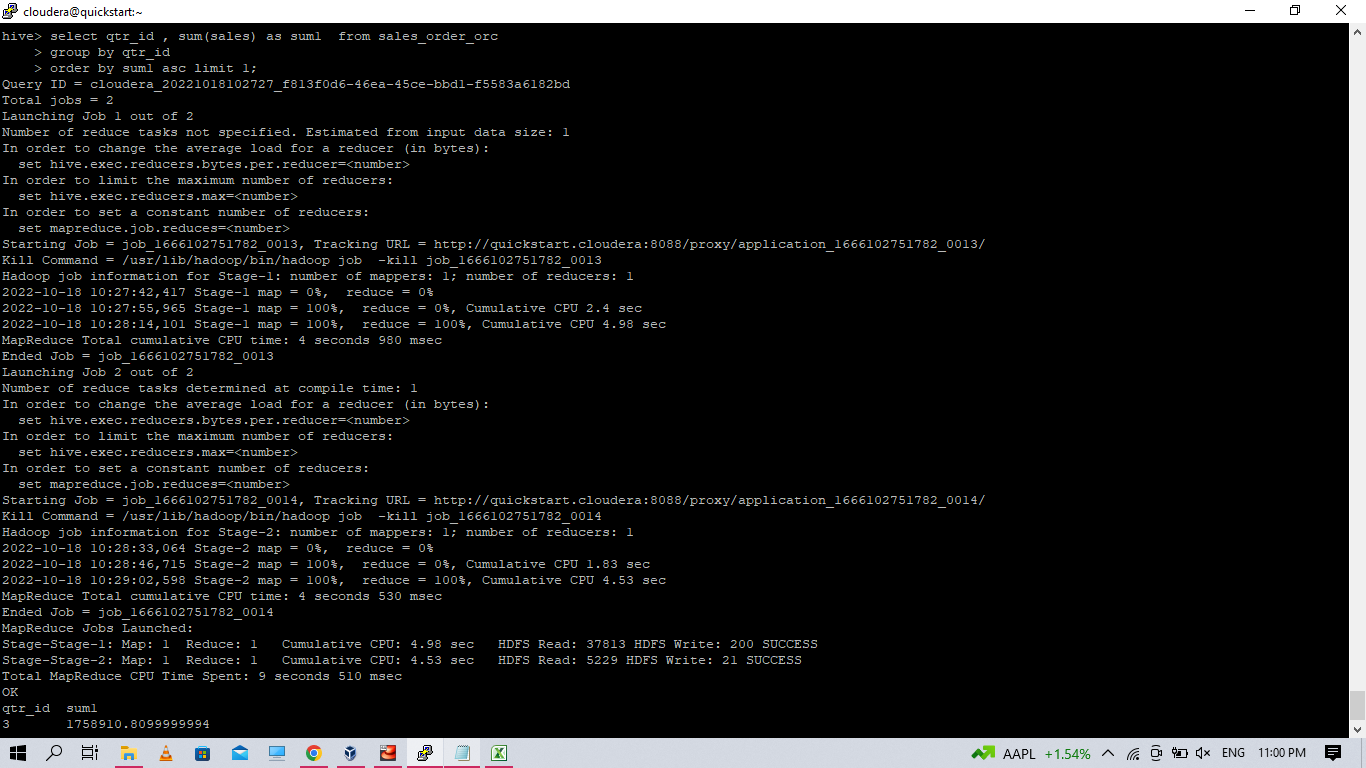


1. **In which quarter sales was minimum**

select qtr\_id , sum(sales) as sum1 from sales\_order\_orc

group by qtr\_id

order by sum1 asc limit 1;



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

select country ,sum(sales) as sum1 , "min sales" from sales\_order\_orc

group by country

order by sum1

limit 1

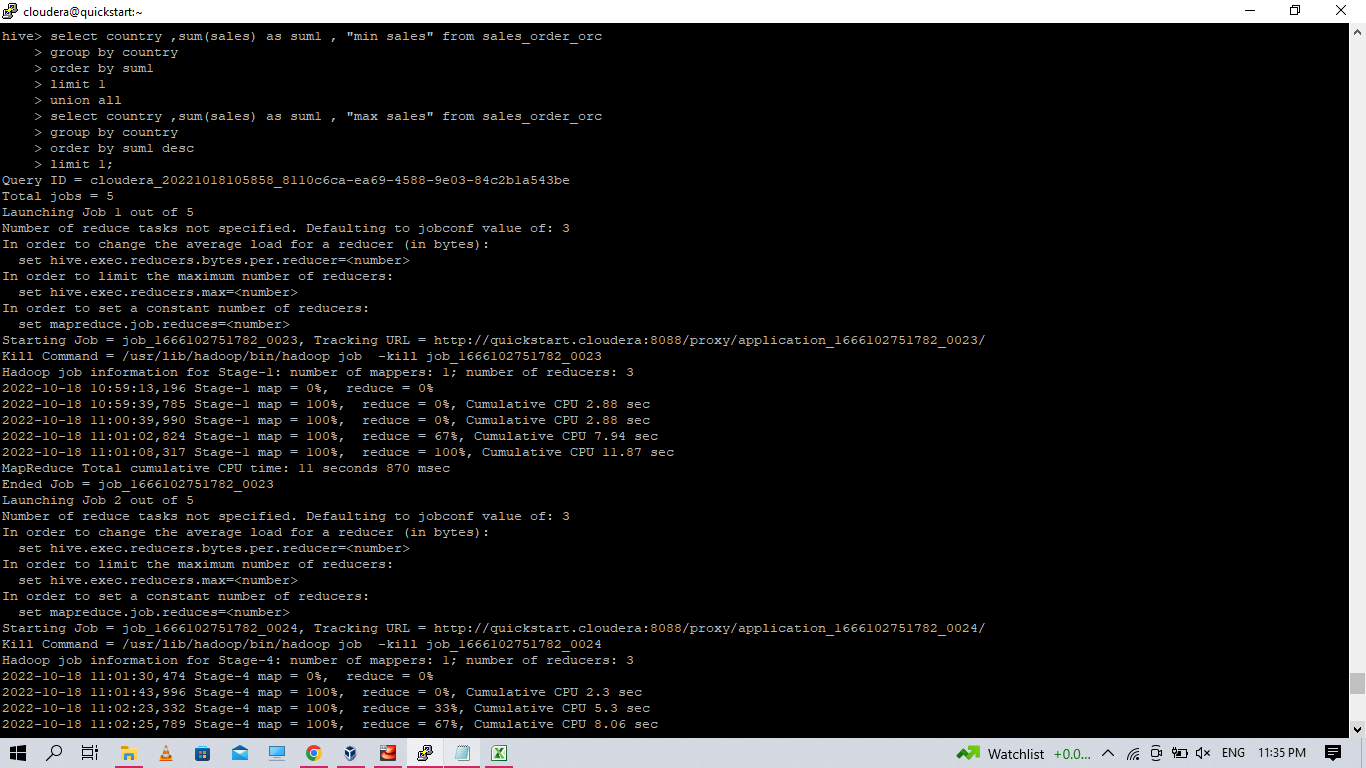
union all

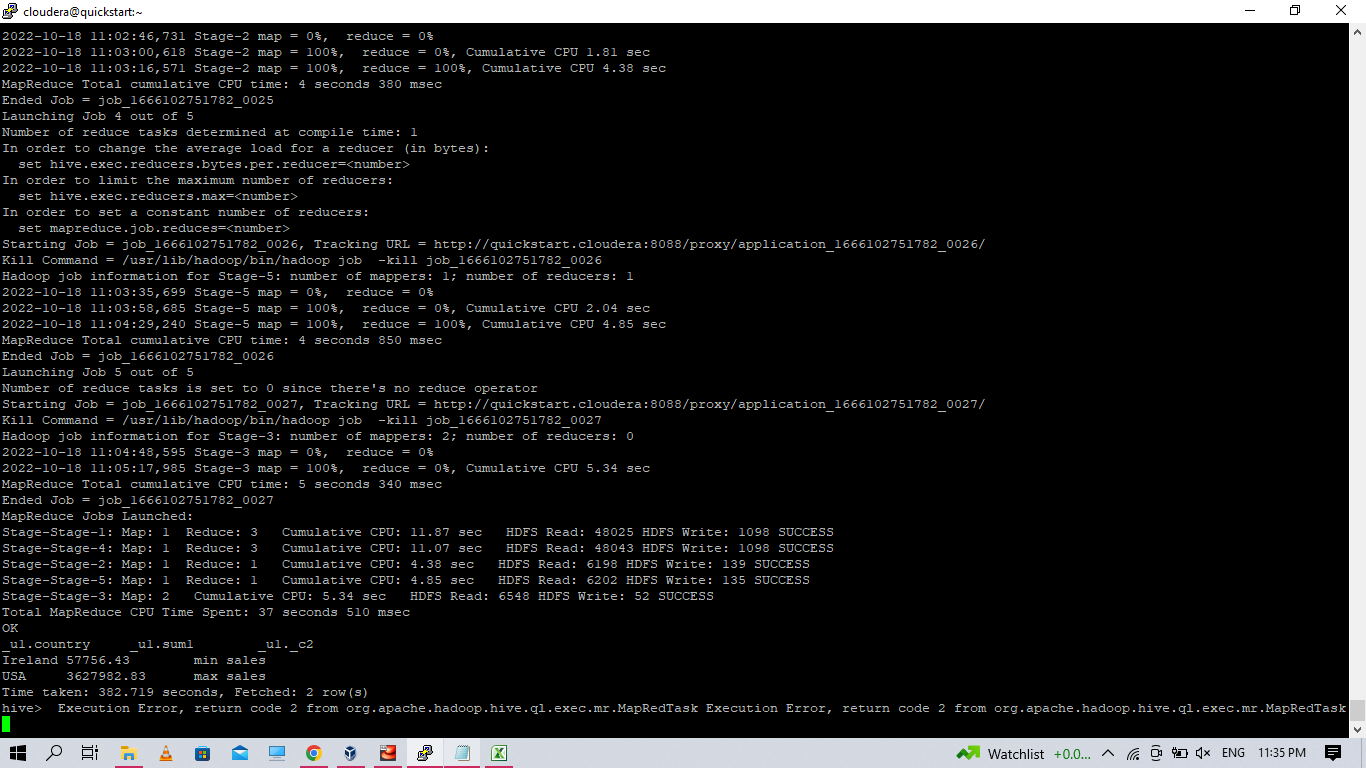
select country ,sum(sales) as sum1 , "max sales" from sales\_order\_orc

group by country

order by sum1 desc

limit 1;



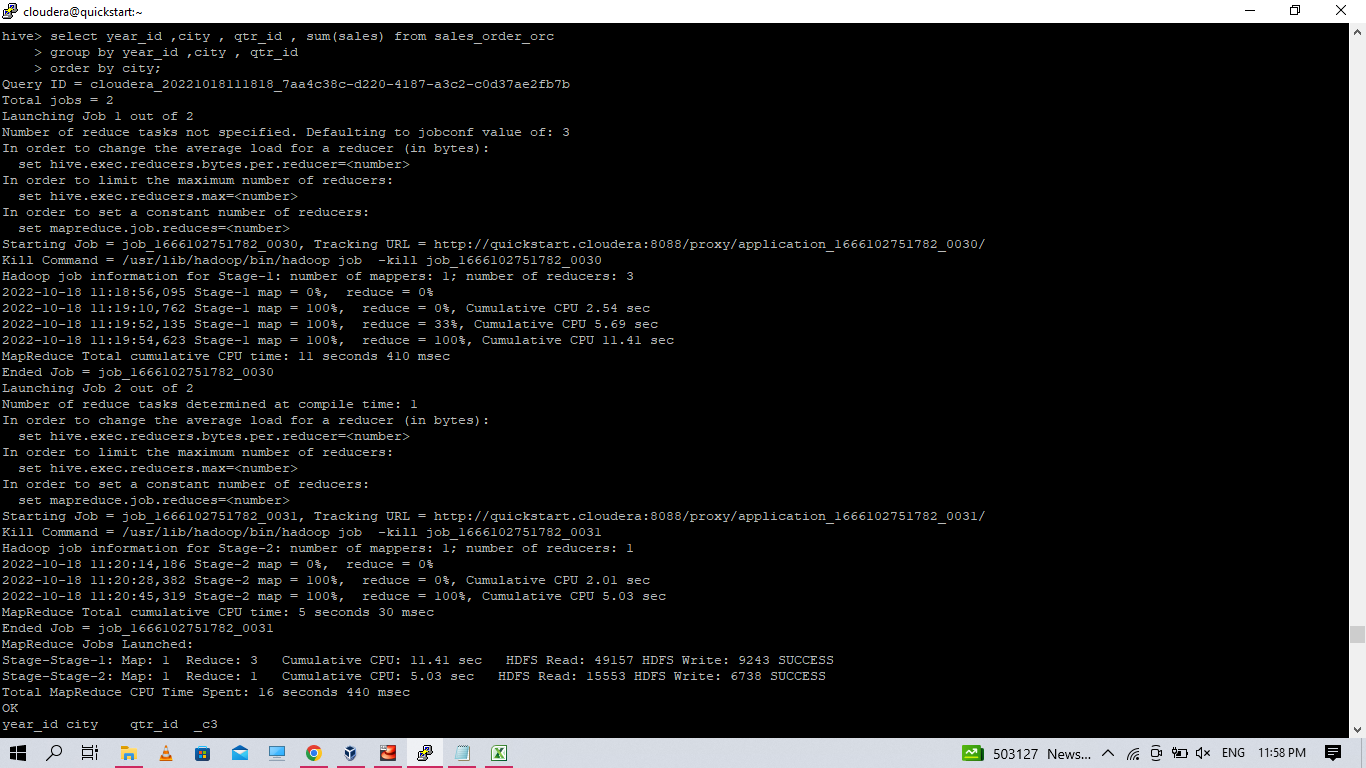


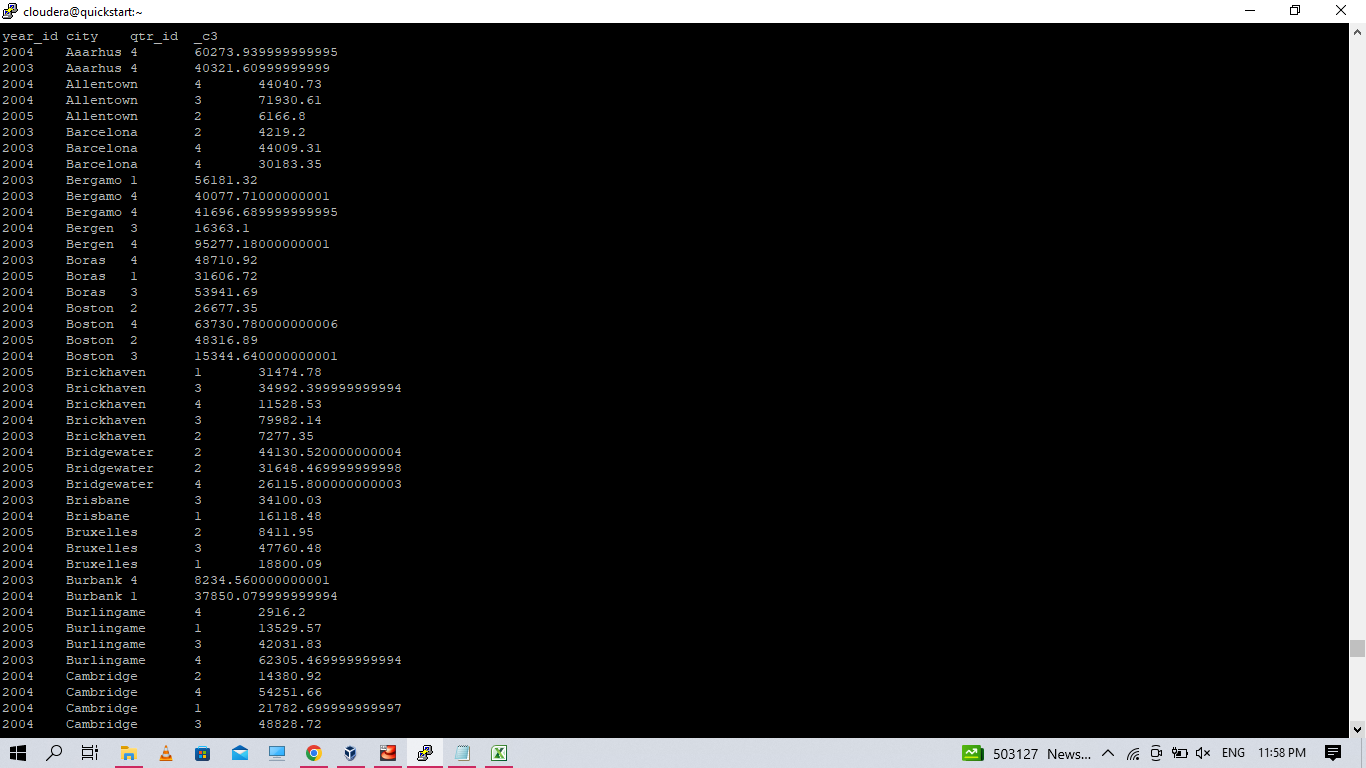
1. **Calculate quartelry sales for each city**

select year\_id , city , qtr\_id , sum(sales) from sales\_order\_orc

group by year\_id , city , qtr\_id

order by city;





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

select \* from

(

select \*, rank() over (partition by sub2.year\_id order by sub2.sum1 desc) as rank

from

(

select year\_id, month\_id, round(max(sales),2) as sum1 from sales\_order\_orc

group by year\_id, month\_id

order by sum1

) sub2

) sub1

where sub1.rank = 1;

