Hive\_Class\_3\_Assignment

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

**-->sales\_order\_data.csv**

**In local path - /home/cloudera/sales\_order\_data.csv**

**Using Filezilla**

2. Store raw data into hdfs location

**[cloudera@quickstart ~]$ hdfs dfs -put sales\_order\_data.csv /challenge**

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

**hive> create table sales\_order\_data\_v1**

**(**

**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”);**

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

**hive> load data inpath '/challenge/sales\_order\_data.csv' into table sales\_order\_data\_v1;**

**hive> set hive.cli.print.header = true;**

**( property to show headers )**

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

**hive> 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;**

6. Load data from "sales\_order\_data\_v1" into "sales\_order\_orc"

**hive> from sales\_order\_csv insert overwrite table sales\_order\_orc select \*;**

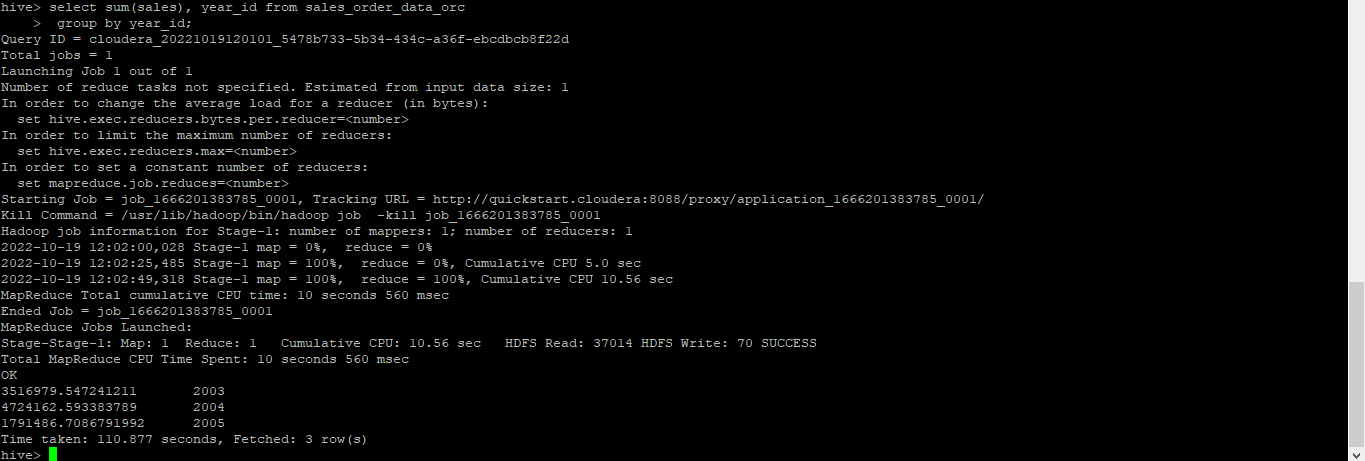
**or**

**hive> insert overwrite table sales\_order\_orc select \* from sales\_order\_data\_v1;**

**a.** Calculatye total sales per year

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

**group by year\_id;**



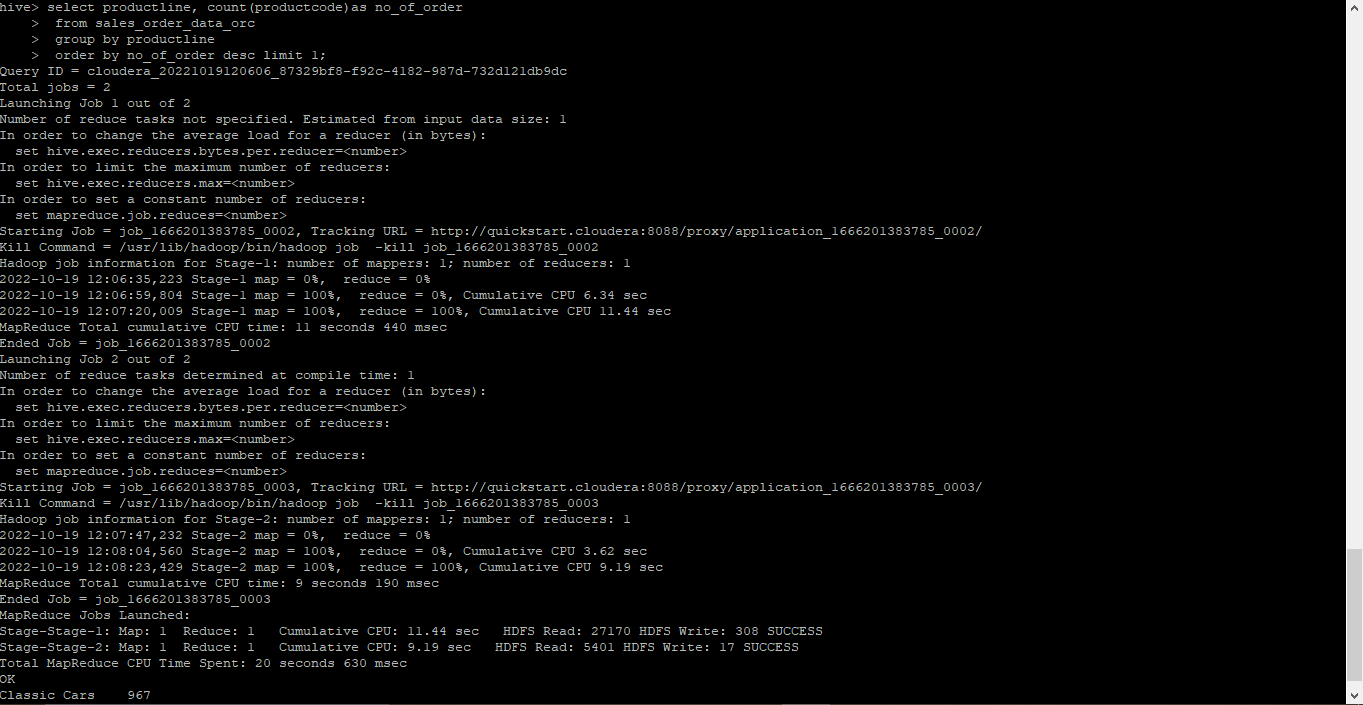
**b.** Find a product for which maximum orders were placed

**select productline, count(productcode)as no\_of\_order**

**from sales\_order\_data\_orc**

**group by productline**

**order by no\_of\_order desc limit 1;**

****

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

**select qtr\_id, sum(sales) as total\_sales**

**from sales\_order\_data\_orc**

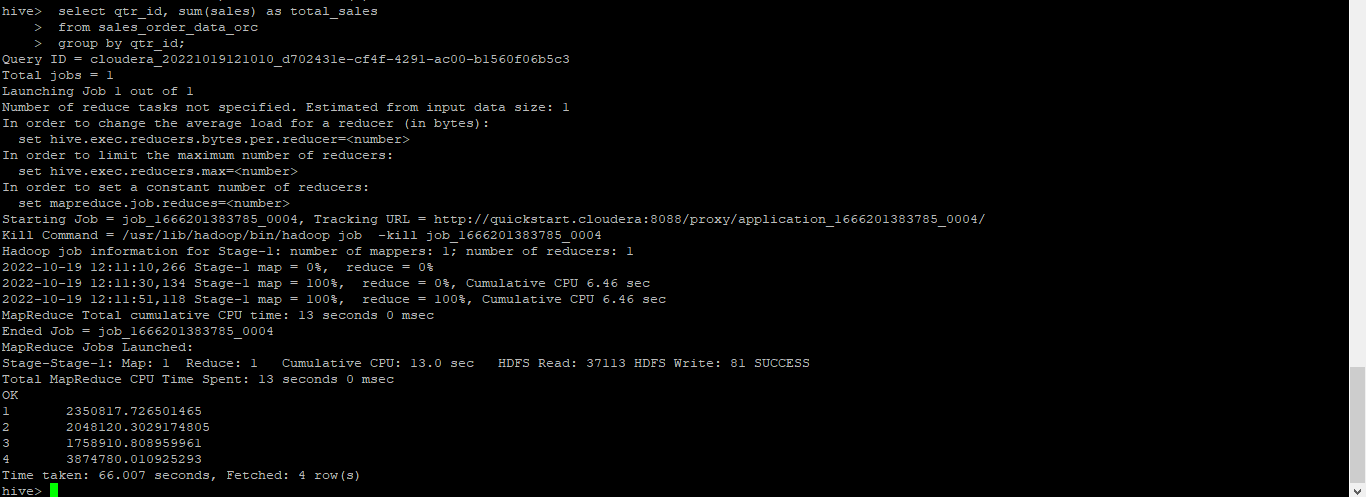
**group by qtr\_id;**

**(for year wise quater sales)**

**select year\_id, qtr\_id, sum(sales) as total\_sales**

**from sales\_order\_data\_orc**

**group by year\_id,qtr\_id;**

****

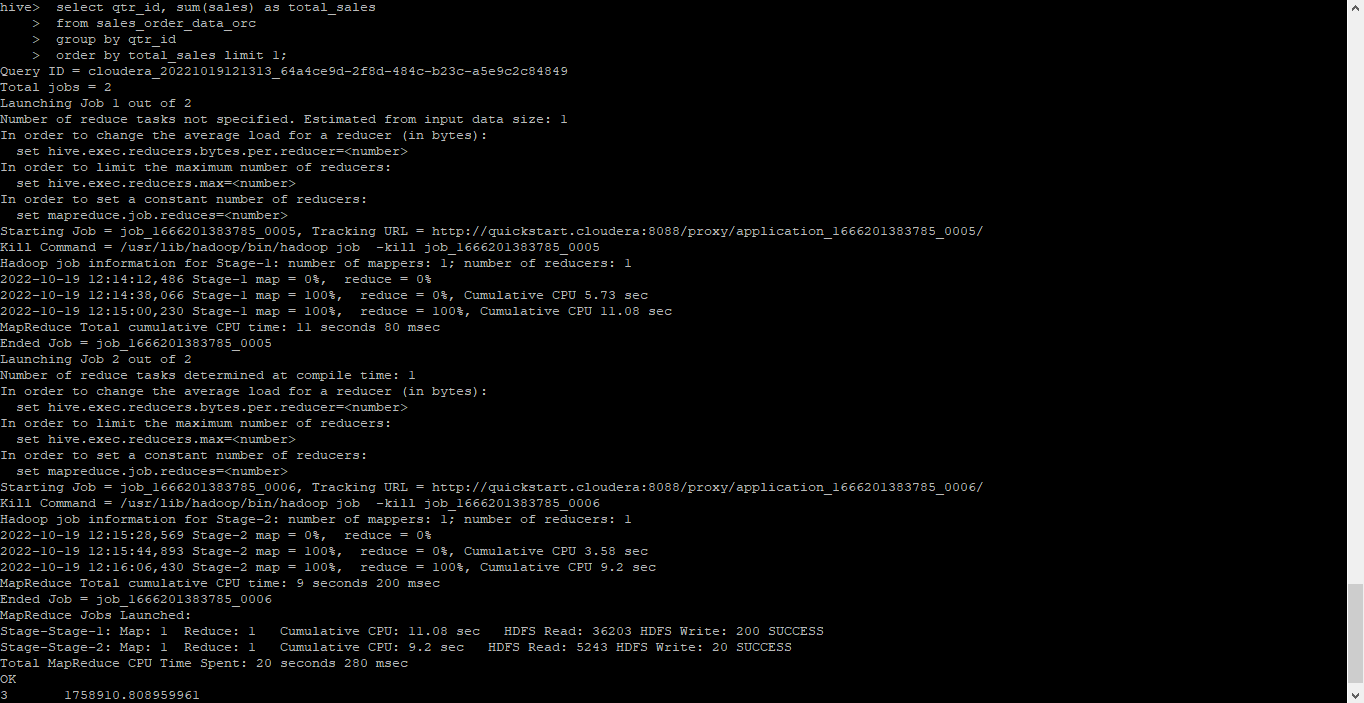
**d.** In which quarter sales was minimum

**select qtr\_id, sum(sales) as total\_sales**

**from sales\_order\_data\_orc**

**group by qtr\_id**

**order by total\_sales limit 1;**

****

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

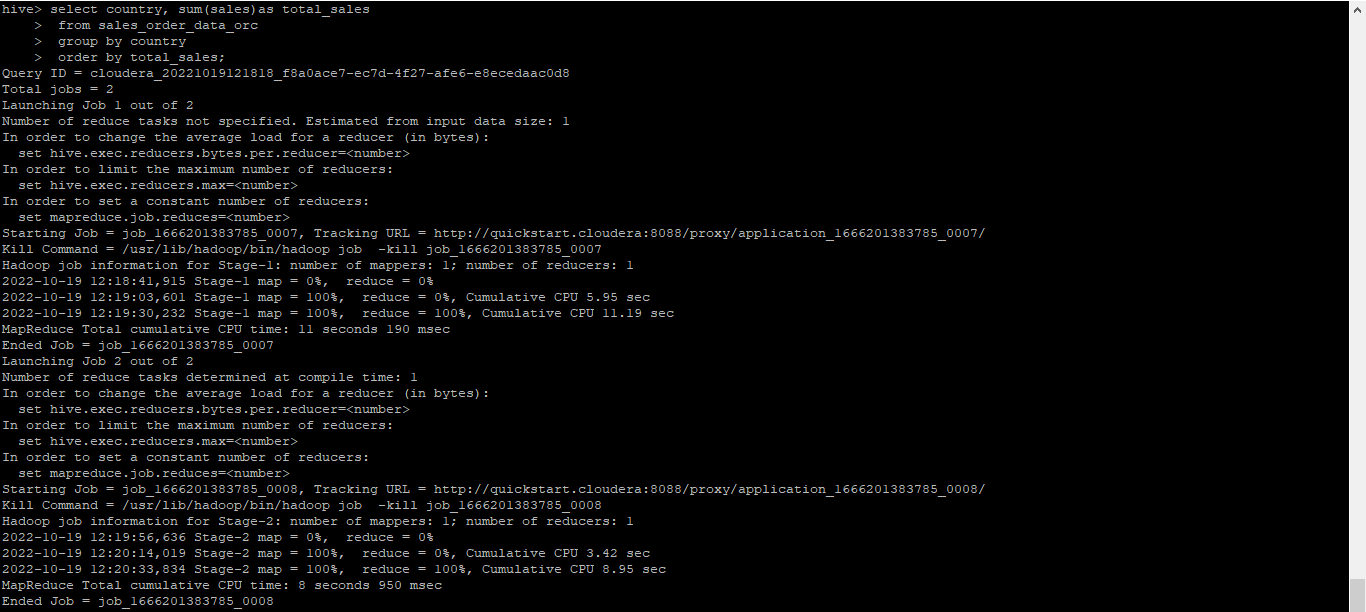
**For min sales :**

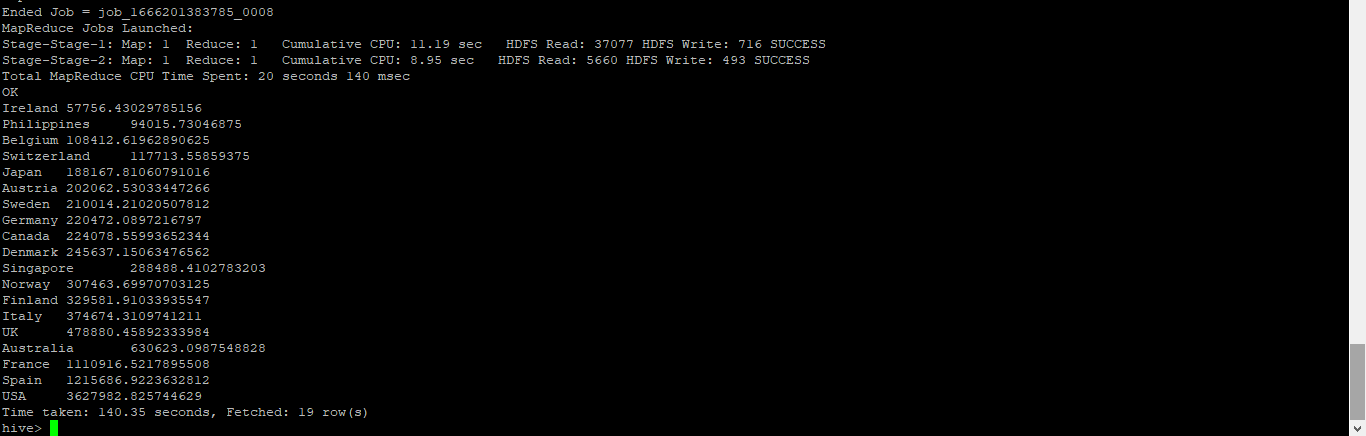
**select country, sum(sales)as total\_sales**

**from sales\_order\_data\_orc**

**group by country**

**order by total\_sales;**

****

****

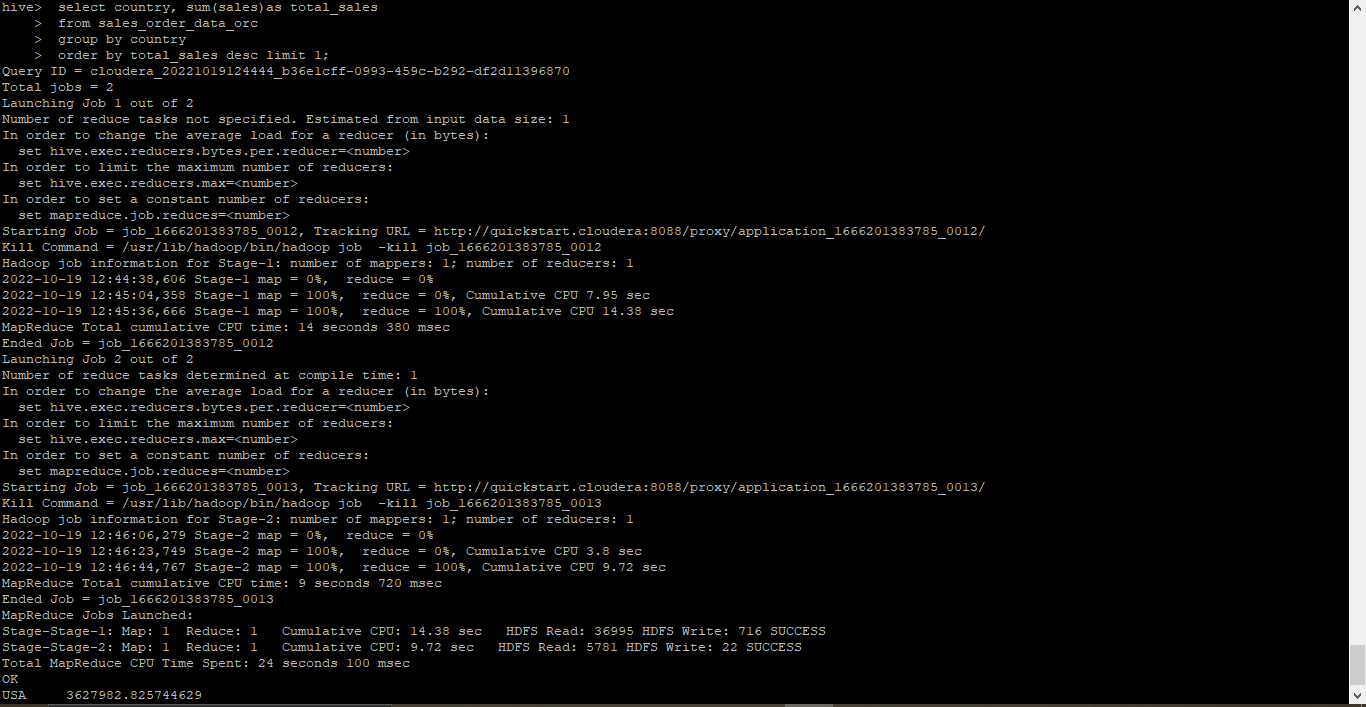
**For max sales :**

**select country, sum(sales)as total\_sales**

**from sales\_order\_data\_orc**

**group by country**

**order by total\_sales desc limit 1;**

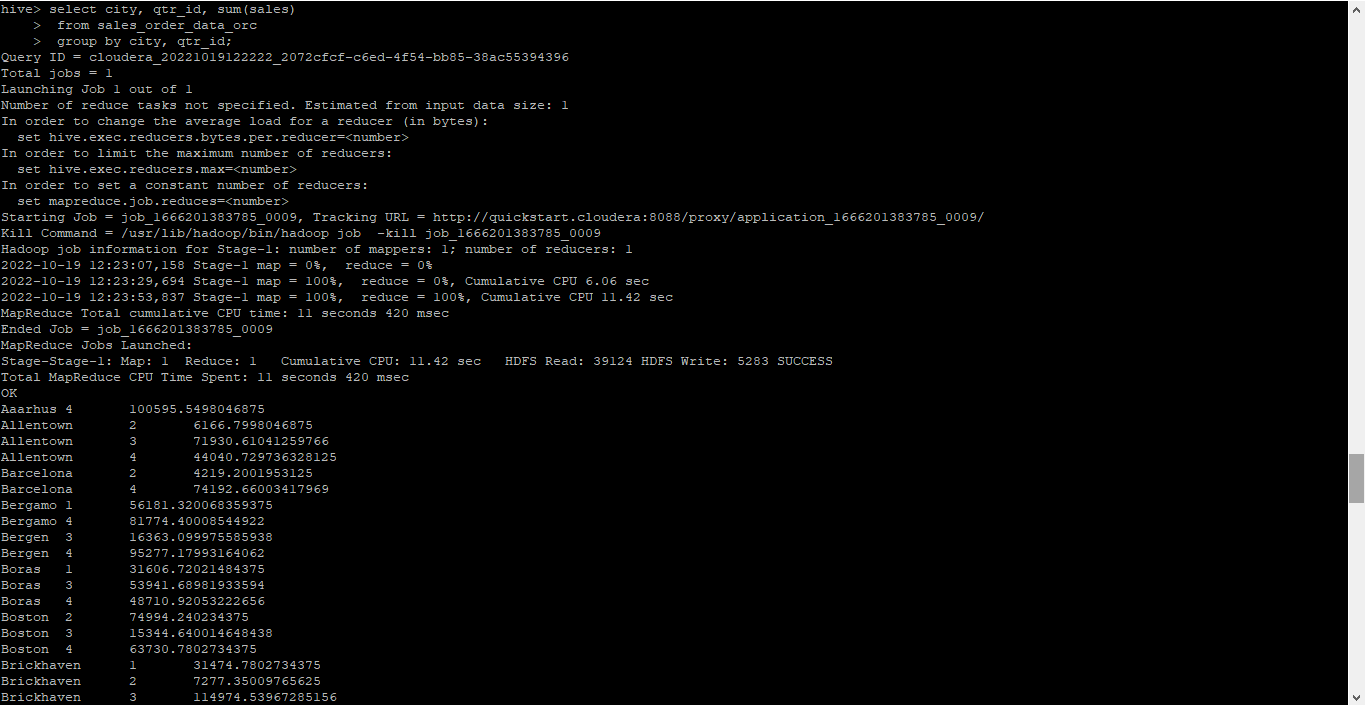
****

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

**select city, qtr\_id, sum(sales)**

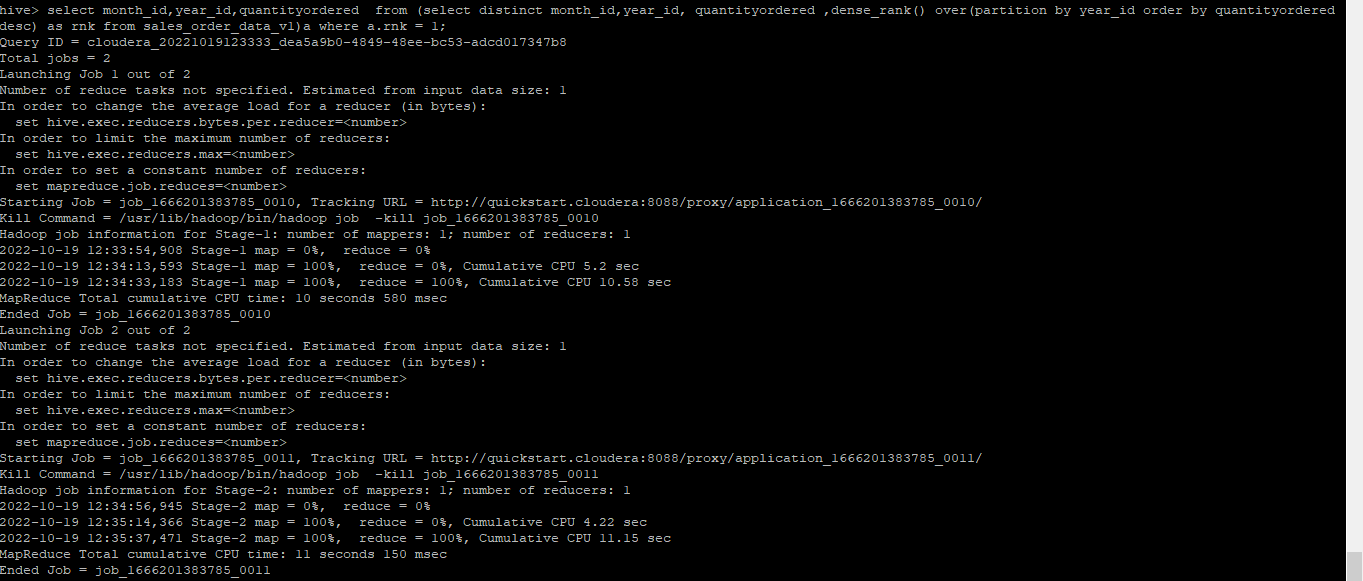
**from sales\_order\_data\_orc**

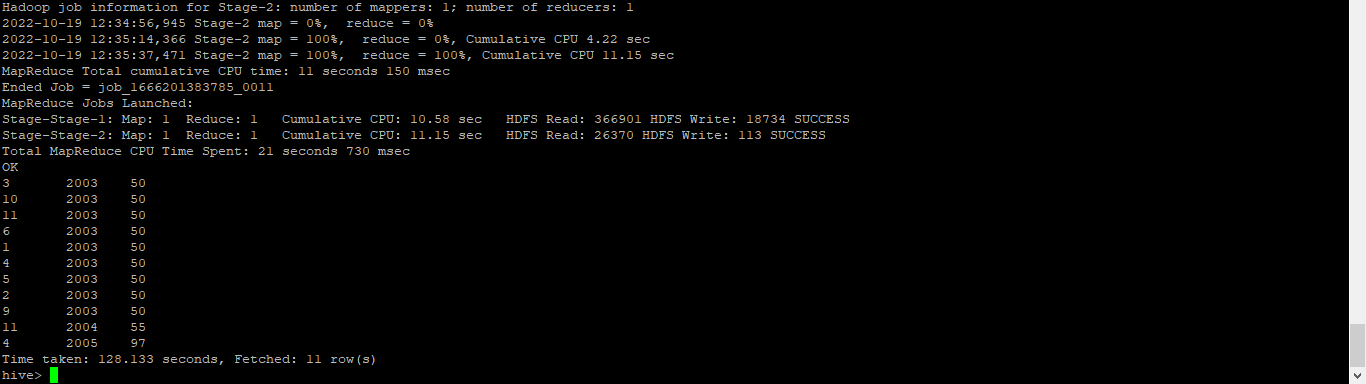
**group by city, qtr\_id;**

****

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

**select month\_id,year\_id,quantityordered from (select distinct month\_id,year\_id, quantityordered ,dense\_rank() over(partition by year\_id order by quantityordered desc) as rnk from sales\_order\_dat\_v1)a where a.rnk = 1;**

****

****