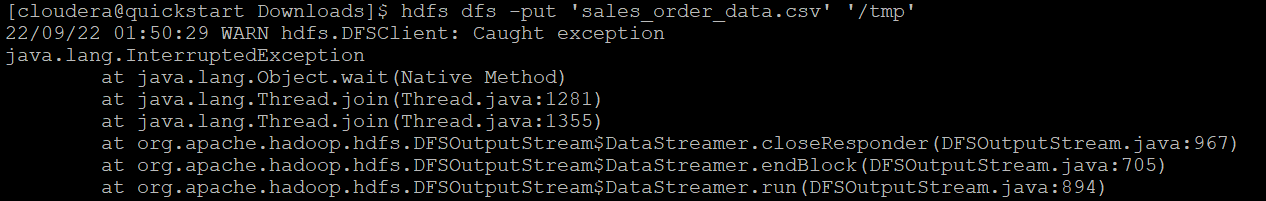
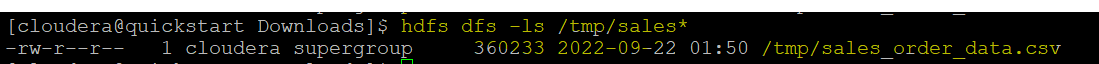
**1**. Download vechile sales data -> <https://github.com/shashank-mishra219/Hive-Class/blob/main/sales_order_data.csv>



**2**. Store raw data into hdfs location

**Command** – hdfs dfs -put ‘sales\_order\_data.csv’ ‘/tmp’





**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

**Command -**

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,

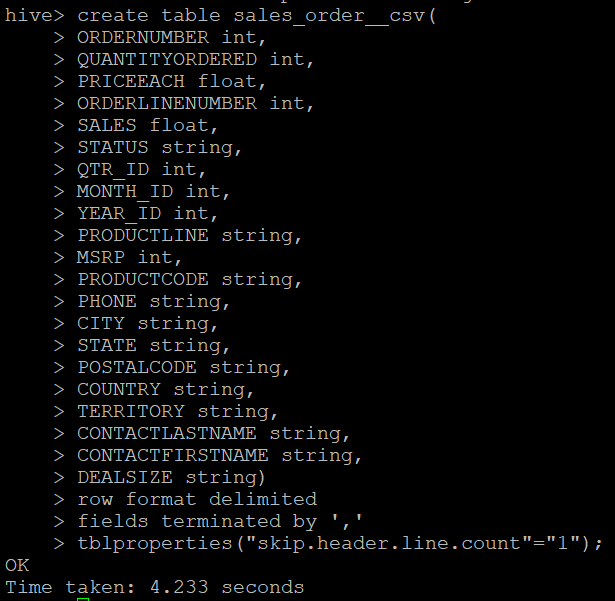
DEALSIZE string

)

row format delimited

fields terminated by ','

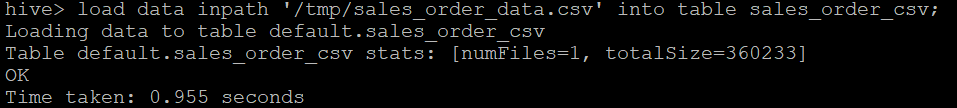
tblproperties("skip.header.line.count"="1");

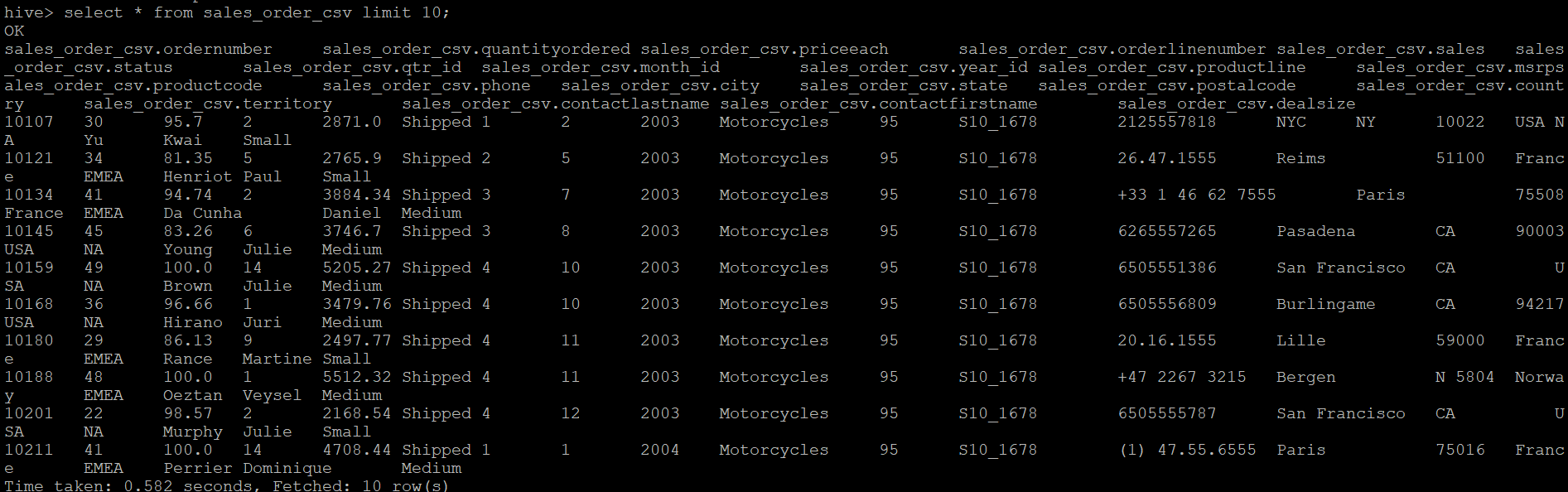


**4.** Load data from hdfs path into "sales\_order\_csv"

**Command –**

load data inpath ‘/tmp/sales\_order\_data.csv’ into table sales\_order\_csv;





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

**Command –**

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,

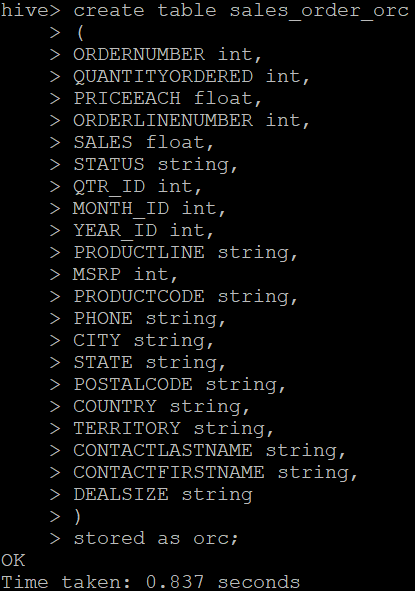
CONTACTLASTNAME string,

CONTACTFIRSTNAME string,

DEALSIZE string

)

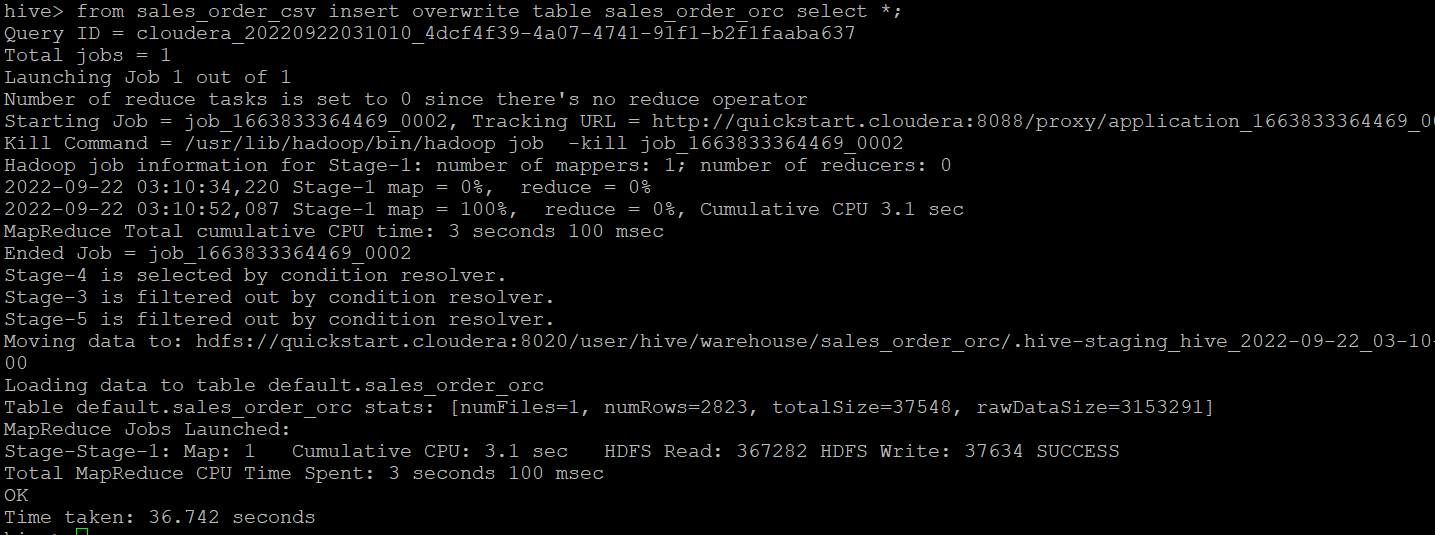
stored as orc;



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

**Command -**

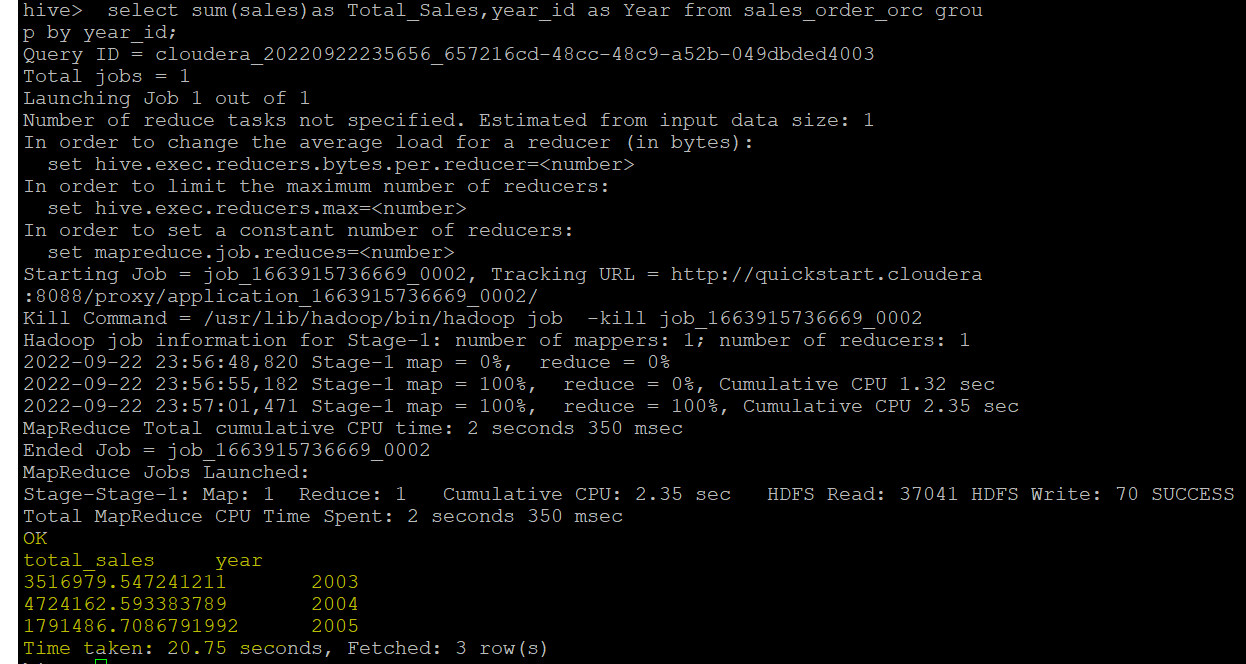
 from sales\_order\_csv insert overwrite table sales\_order\_orc select \*;



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

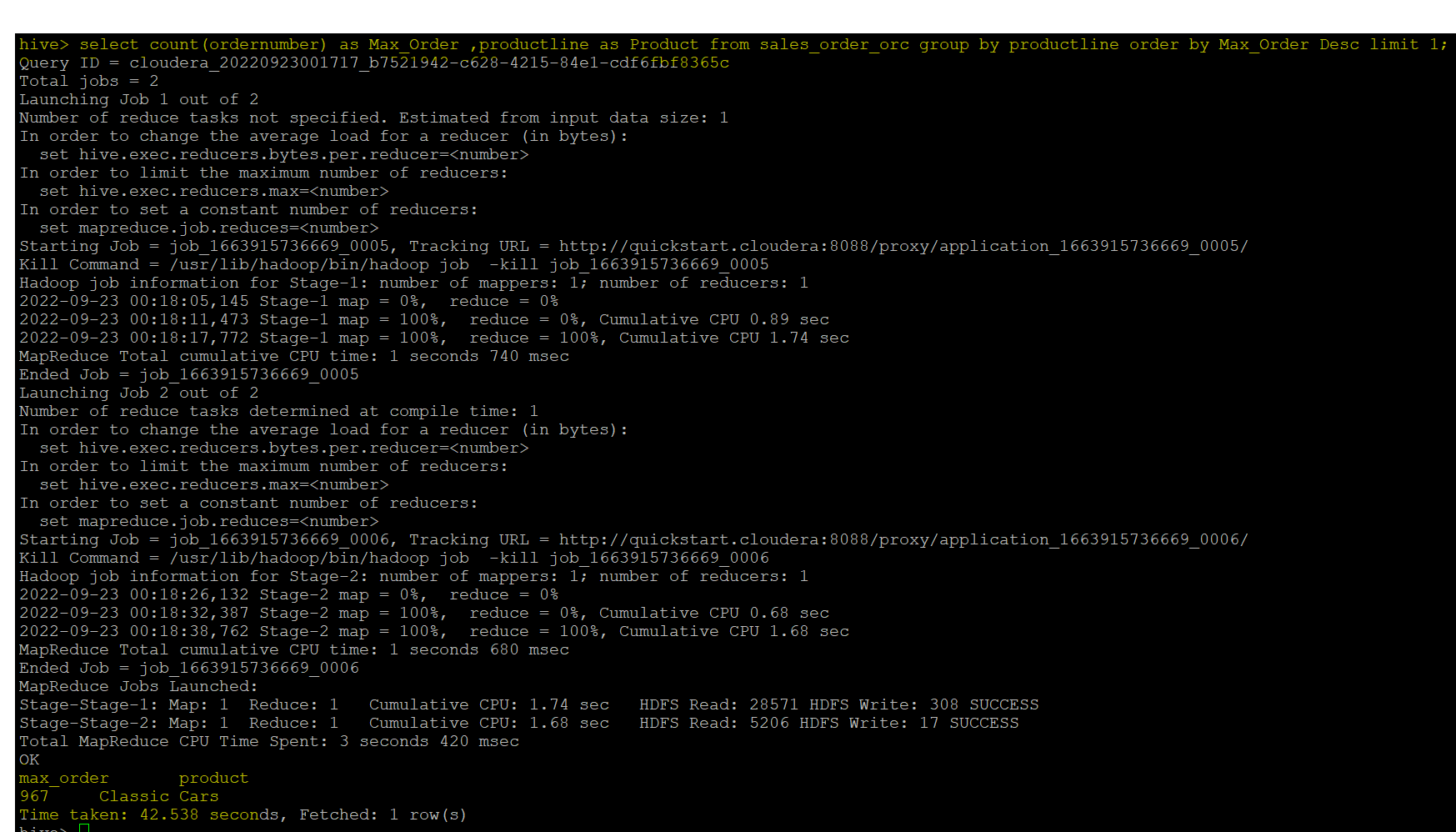
**a. Calculate total sales per year**

select sum(sales)as Total\_Sales,year\_id as Year from sales\_order\_orc group by year\_id;

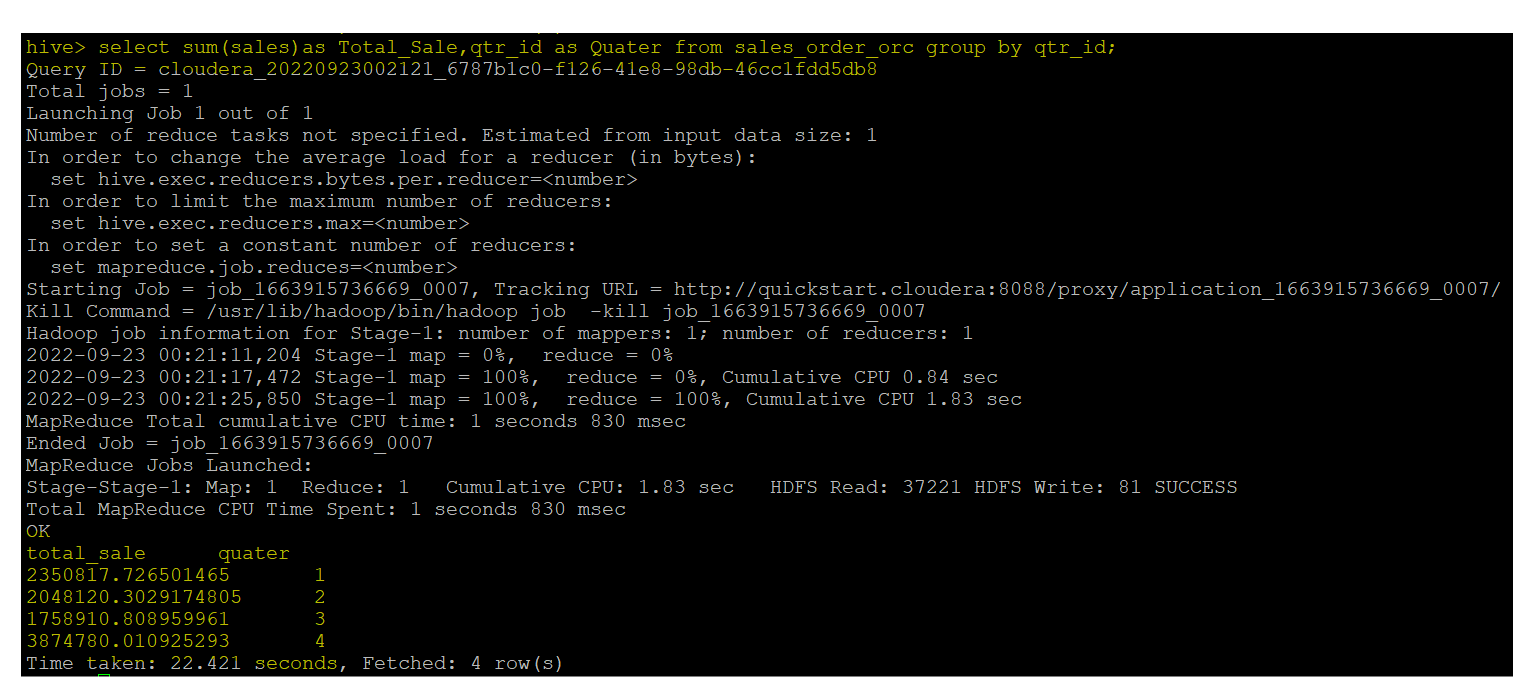


**b. Find a product for which maximum orders were placed –**

select count(ordernumber) as Max\_Order ,productline as Product from sales\_order\_orc group by productline order by Max\_Order Desc limit 1;

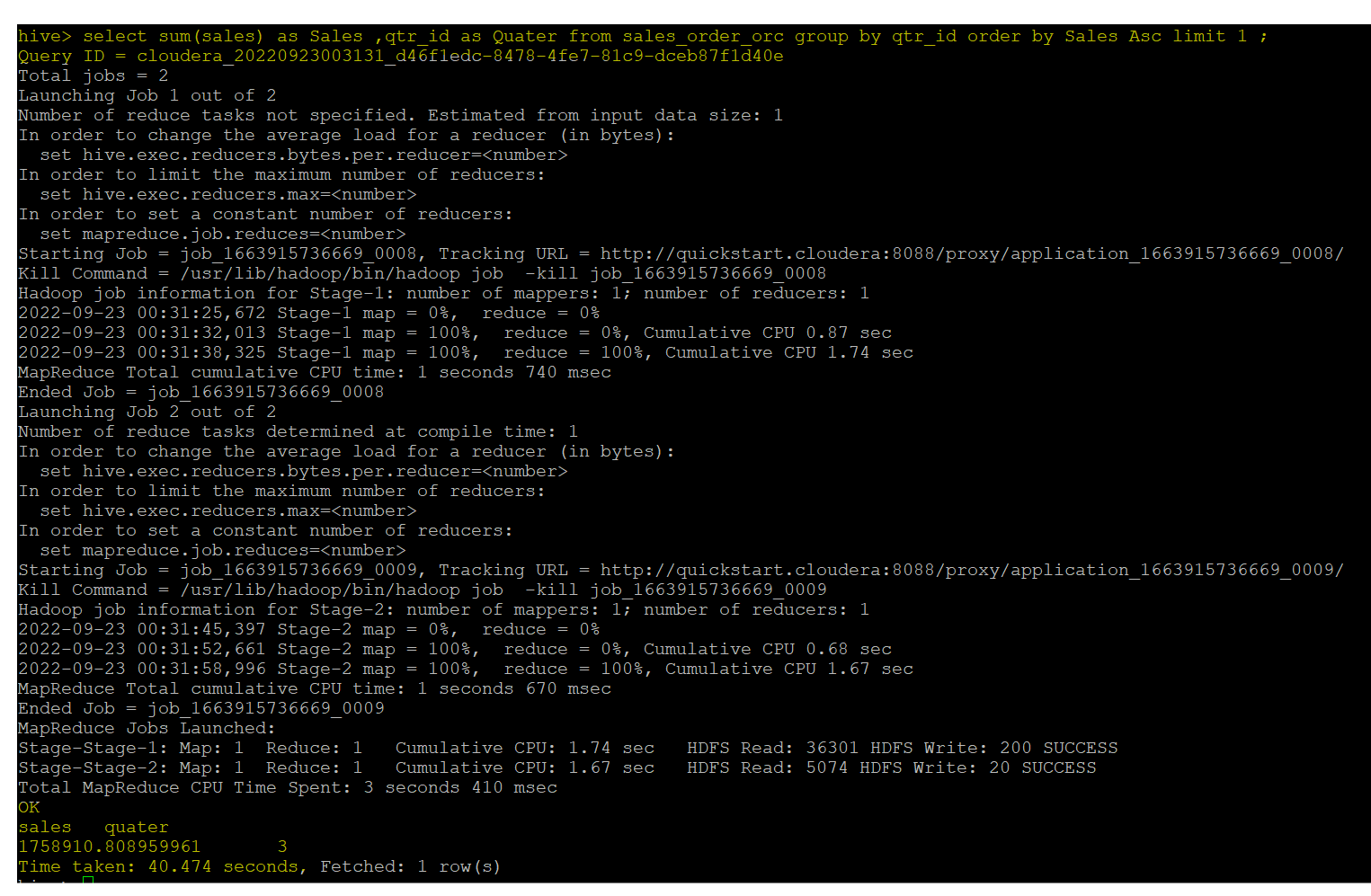


**c. Calculate the total sales for each quarter -**

select sum(sales)as Total\_Sale,qtr\_id as Quater from sales\_order\_orc group by qtr\_id;

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

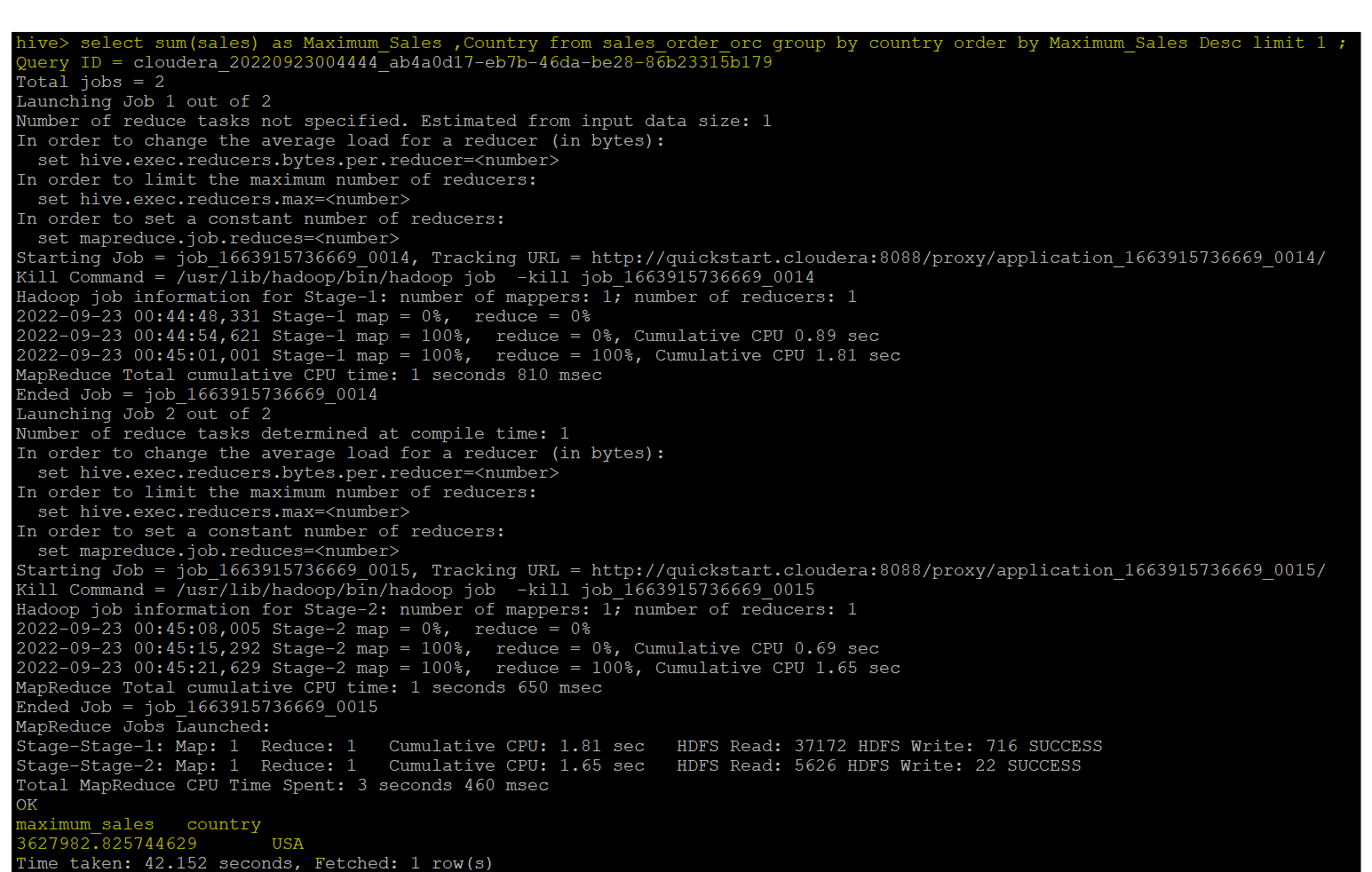
select sum(sales) as Sales ,qtr\_id as Quater from sales\_order\_orc group by qtr\_id order by Sales Asc limit 1 ;



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

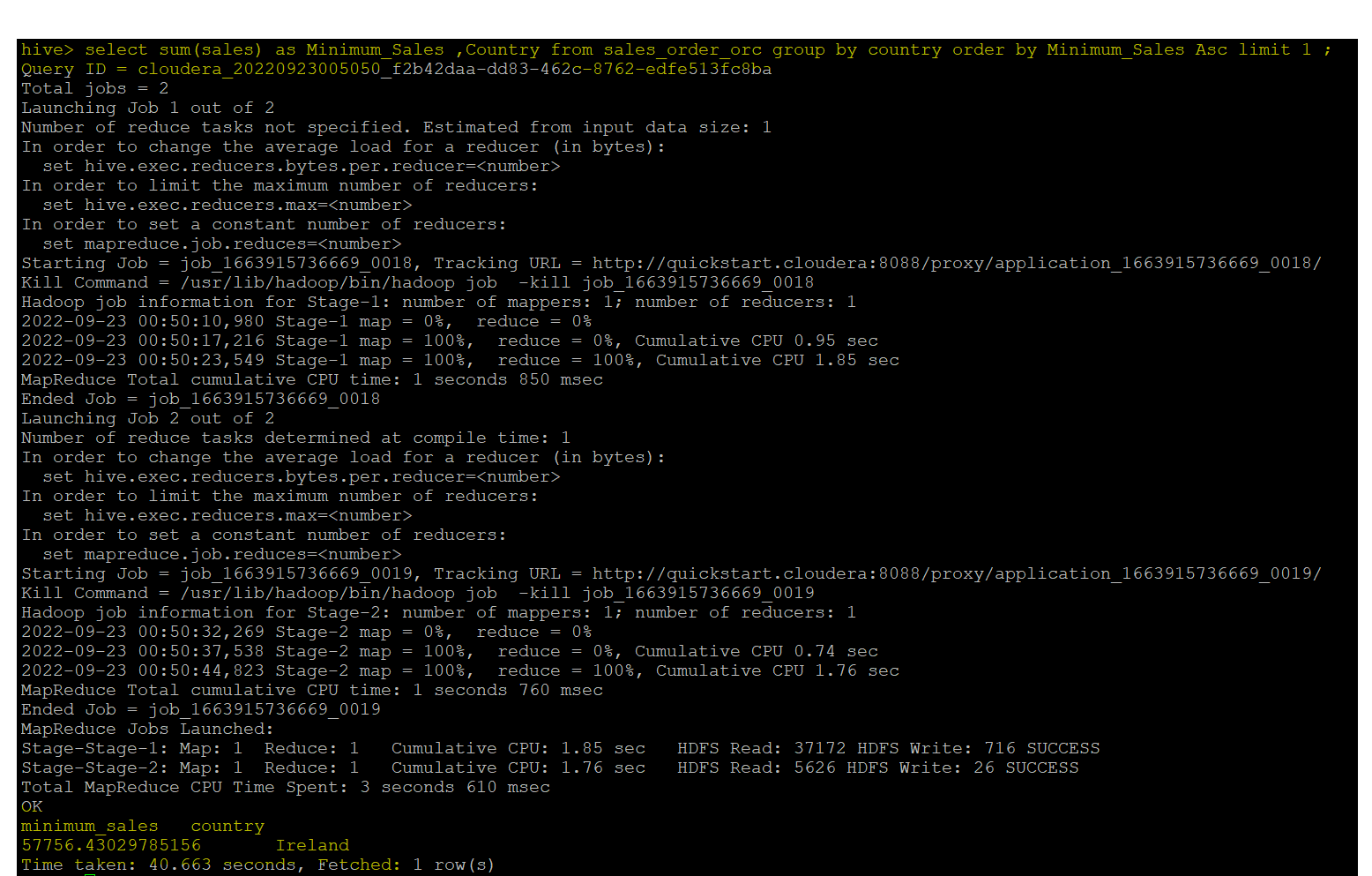
**Maximum -**

select sum(sales) as Maximum\_Sales ,Country from sales\_order\_orc group by country order by Maximum\_Sales Desc limit 1 ;



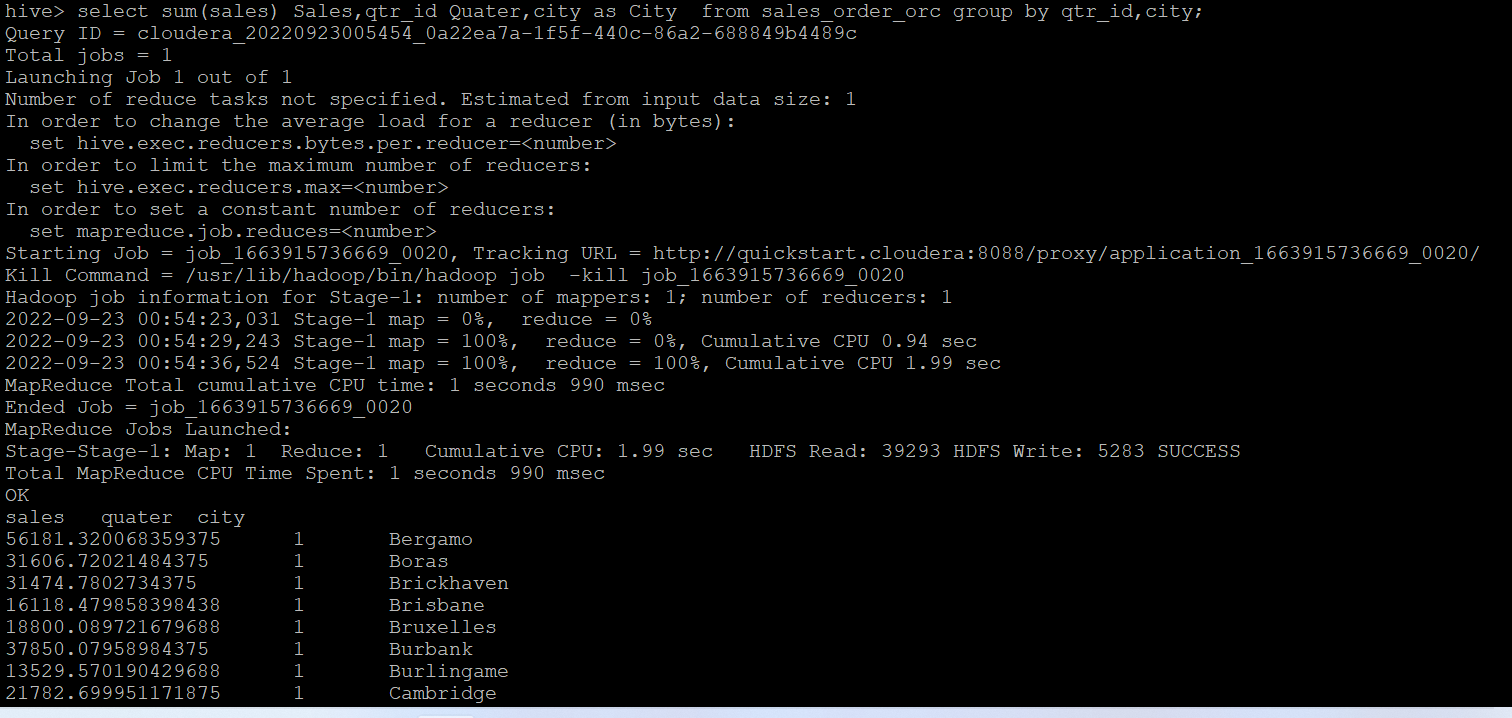
**Minimum Sales –**

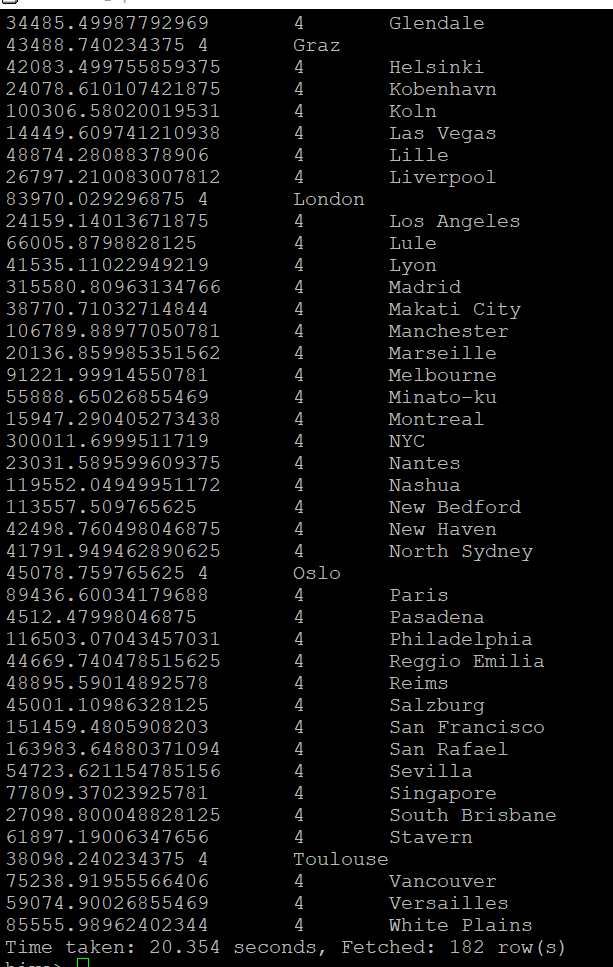
select sum(sales) as Minimum\_Sales ,Country from sales\_order\_orc group by country order by Minimum\_Sales Asc limit 1 ;



**f. Calculate quarterly sales for each city --**

select sum(sales) Sales,qtr\_id Quater,city as City from sales\_order\_orc group by qtr\_id,city;





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

select month\_id, max(QUANTITYORDERED),year\_id from sales\_order\_orc group by year\_id;

