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

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"

create table sales\_order\_csv

(

ORDERNUMBER INT,

QUANTITYORDERED INT,

PRICEEACH INT,

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,

DEALSIZE STRING

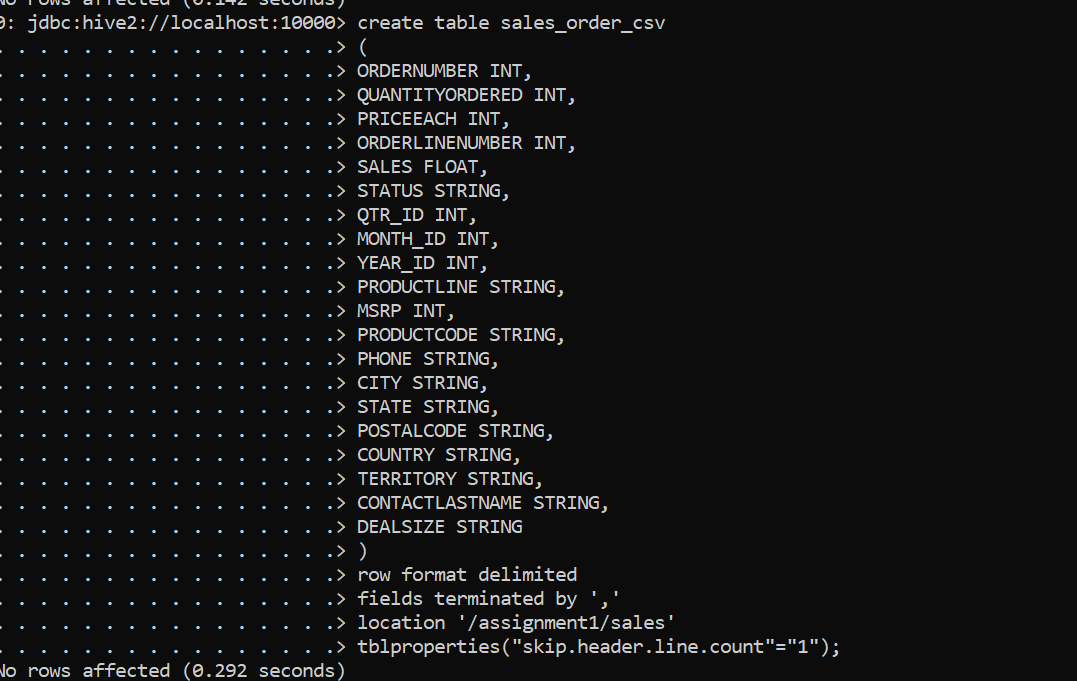
)

row format delimited

fields terminated by ','

location '/assignment1/sales'

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



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

create table sales\_order\_orc

(

ORDERNUMBER INT,

QUANTITYORDERED INT,

PRICEEACH INT,

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,

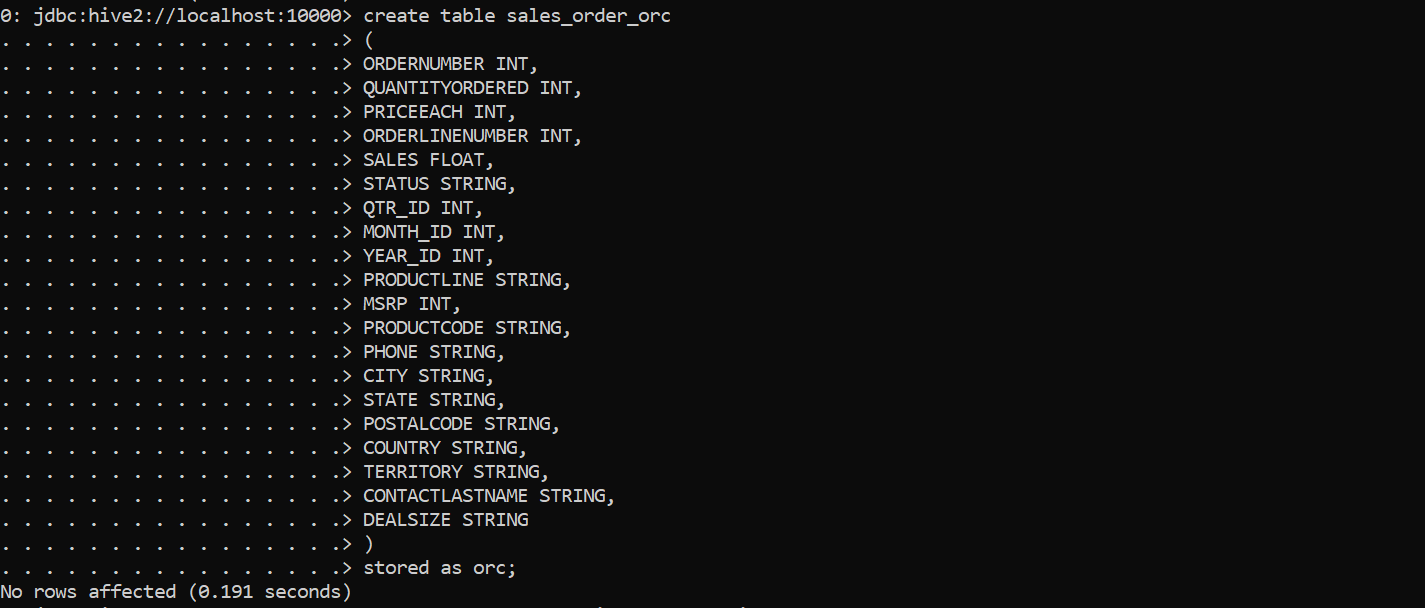
TERRITORY STRING,

CONTACTLASTNAME STRING,

DEALSIZE STRING

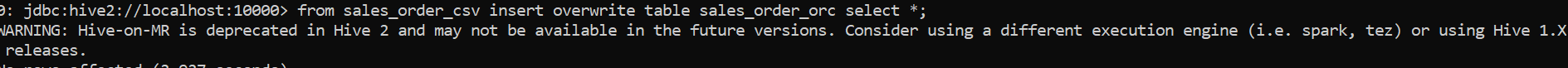
)

stored as orc;



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

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



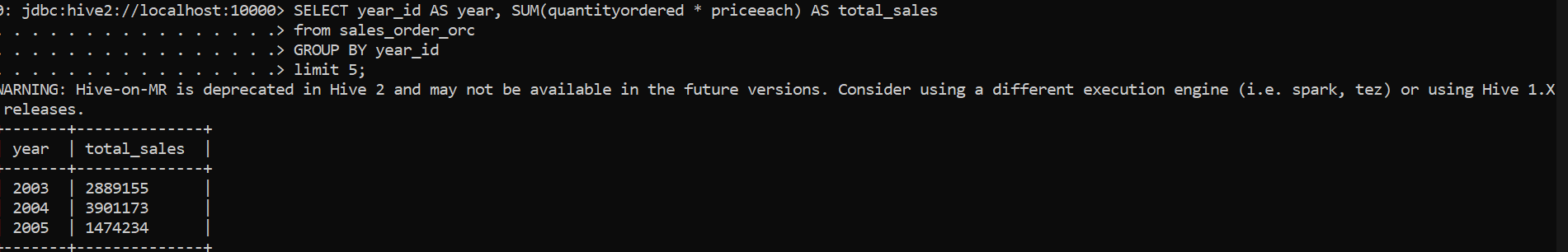
a. Calculate total sales per year

SELECT year\_id AS year, SUM(quantityordered \* priceeach) AS total\_sales

from sales\_order\_orc

GROUP BY year\_id

limit 5;



b. Find a product for which maximum orders were placed

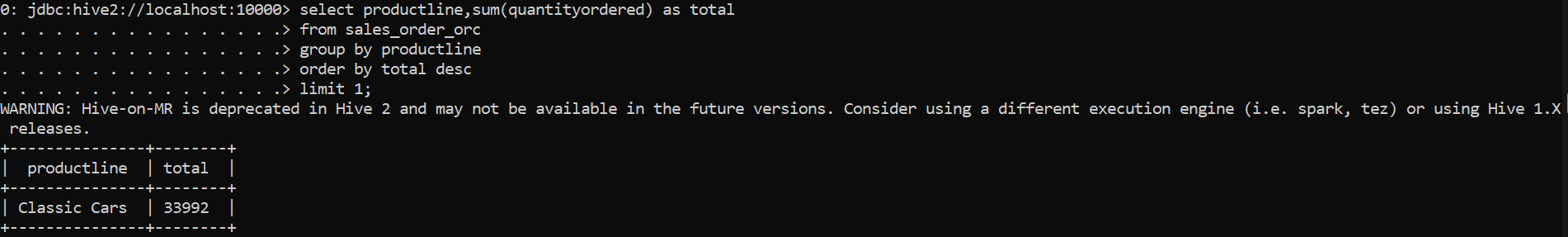
select productline,sum(quantityordered) as total

from sales\_order\_orc

group by productline

order by total desc

limit 1;

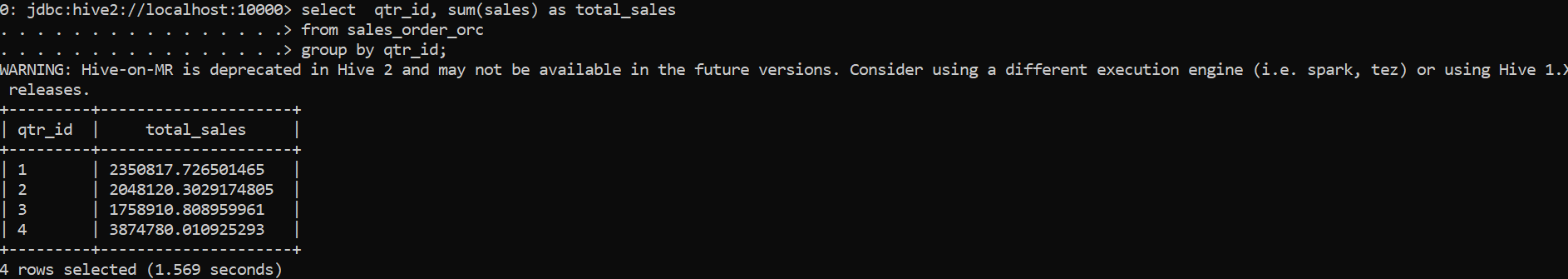


c. Calculate the total sales for each quarter

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

from sales\_order\_orc

group by qtr\_id;



d. In which quarter sales was minimum

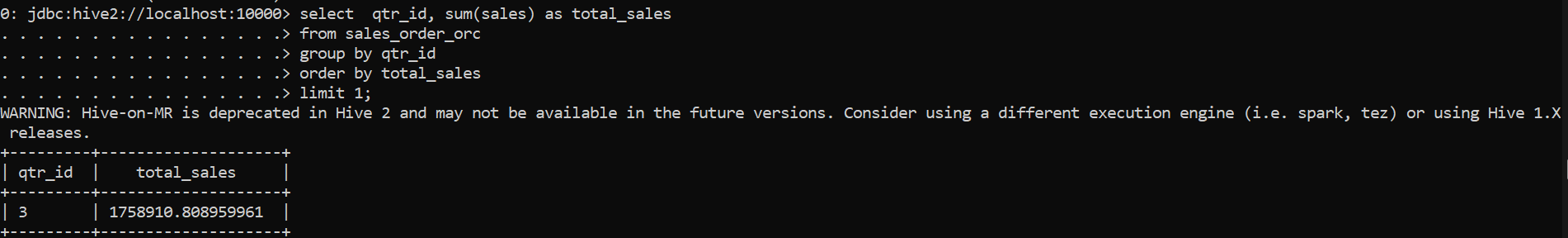
select qtr\_id, sum(sales) as total\_sales

from sales\_order\_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

SELECT

country,

total\_sales

FROM (

SELECT

country,

SUM(sales) as total\_sales,

RANK() OVER (ORDER BY SUM(sales) DESC) as sales\_rank\_desc,

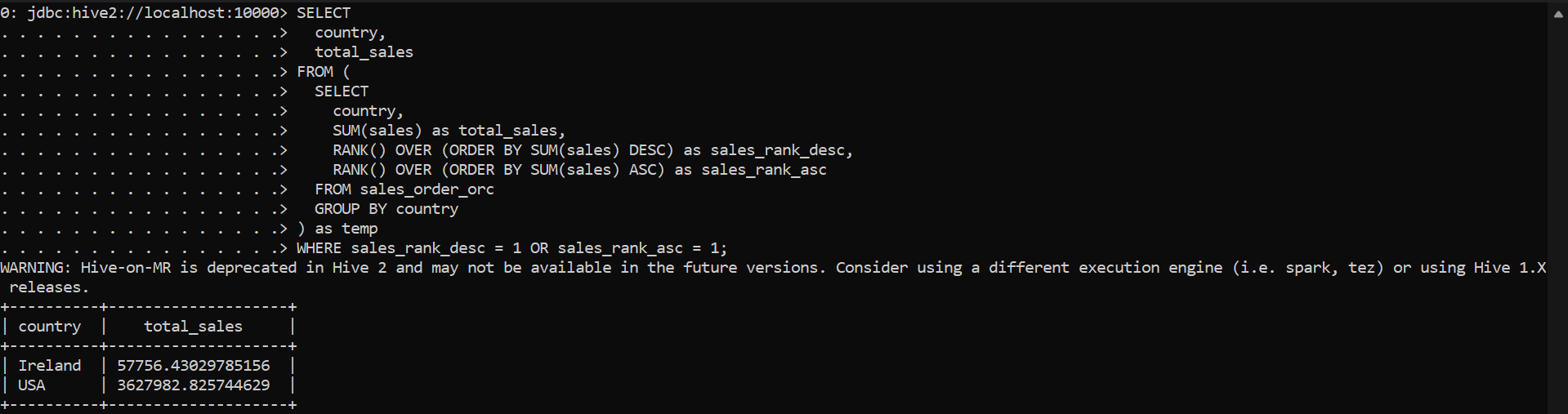
RANK() OVER (ORDER BY SUM(sales) ASC) as sales\_rank\_asc

FROM sales\_order\_orc

GROUP BY country

) as temp

WHERE sales\_rank\_desc = 1 OR sales\_rank\_asc = 1;

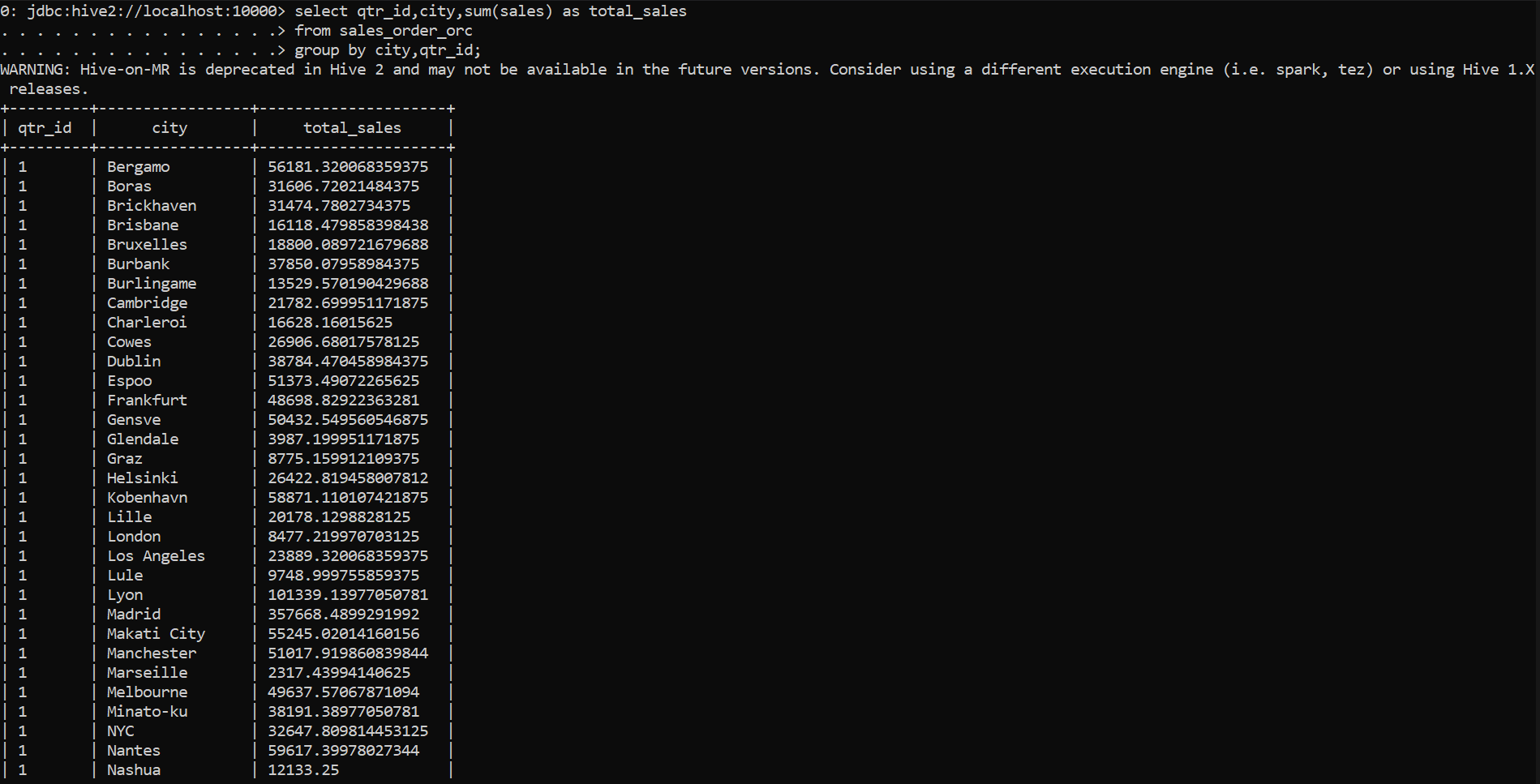


f. Calculate quartelry sales for each city

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

from sales\_order\_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, rnk

FROM (

SELECT month\_id, year\_id, QUANTITYORDERED, RANK() OVER (PARTITION BY year\_id ORDER BY QUANTITYORDERED DESC) AS rnk

FROM sales\_order\_csv

) a

WHERE a.rnk = 1;

