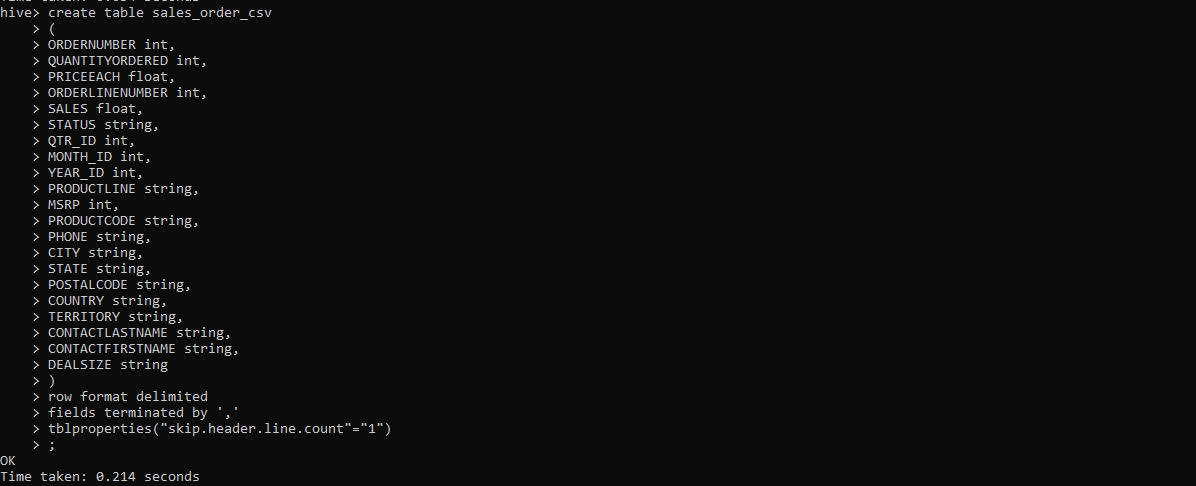
Hive Assignment – 01 solution

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"

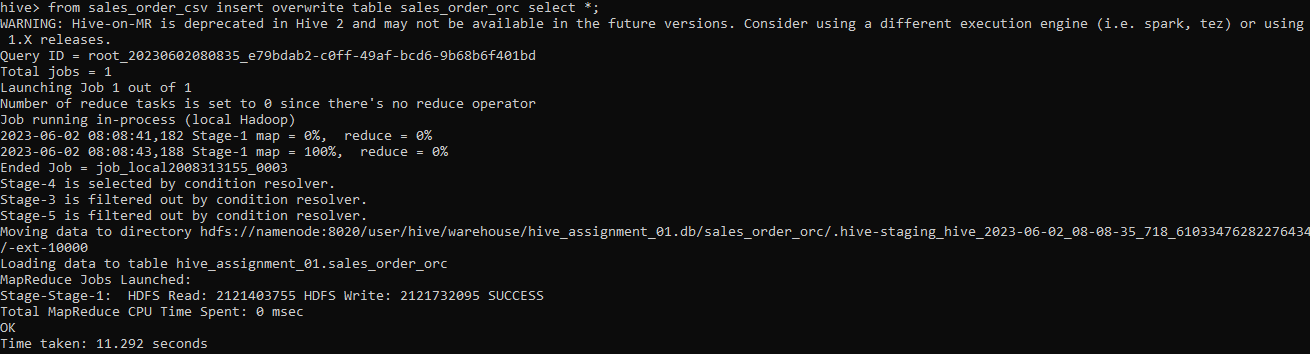
LOAD DATA LOCAL INPATH '/home/sales\_order\_data.csv' INTO TABLE sales\_order\_csv;

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



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

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

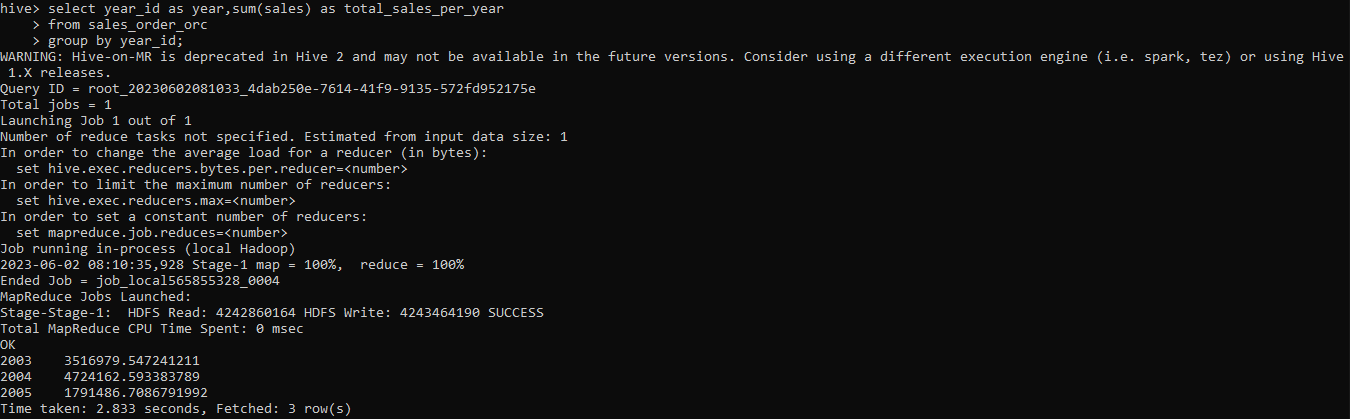


Perform below menioned queries on "sales\_order\_orc" table :

1. Calculate total sales per year
   * select year\_id as year,sum(sales) as total\_sales\_per\_year

from sales\_order\_orc

group by year\_id;



1. Find a product for which maximum orders were placed

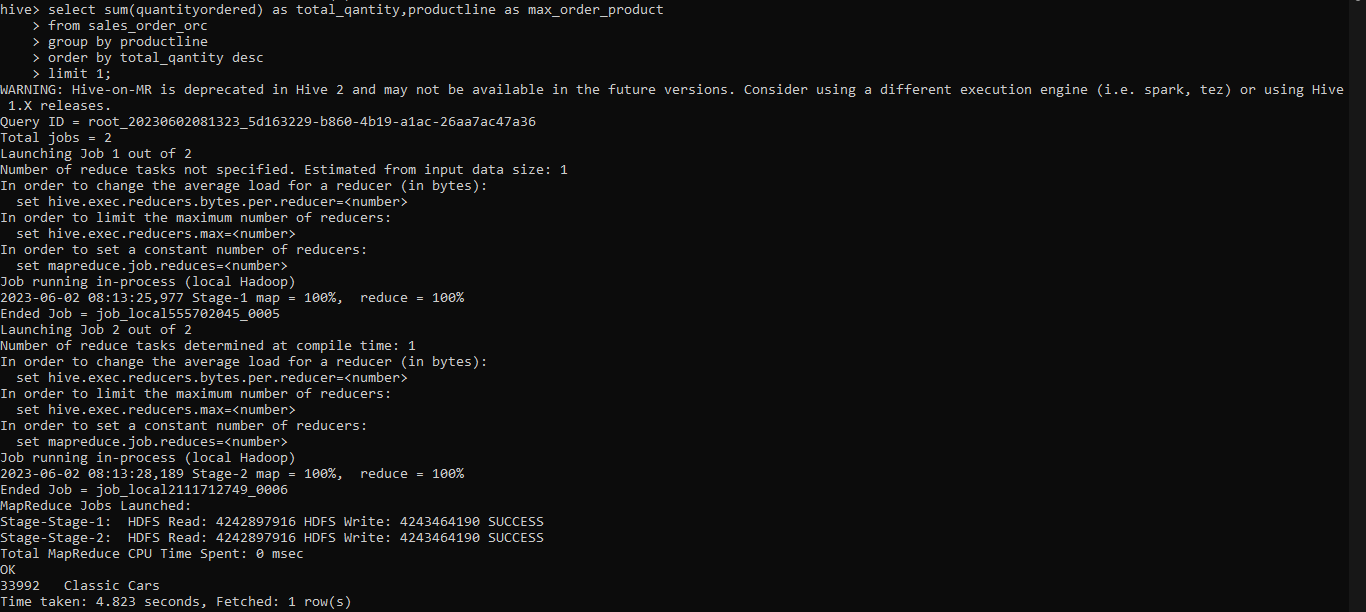
* select sum(quantityordered) as total\_qantity,productline as max\_order\_product

from sales\_order\_orc

group by productline

order by total\_qantity desc

limit 1;

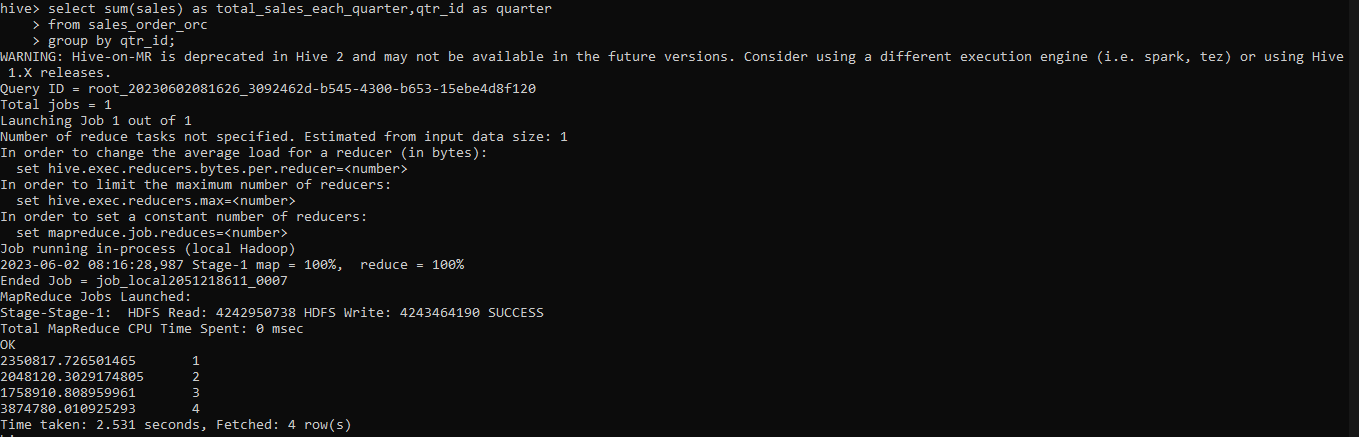


c. Calculate the total sales for each quarter

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

from sales\_order\_orc

group by qtr\_id;



d. In which quarter sales was minimum

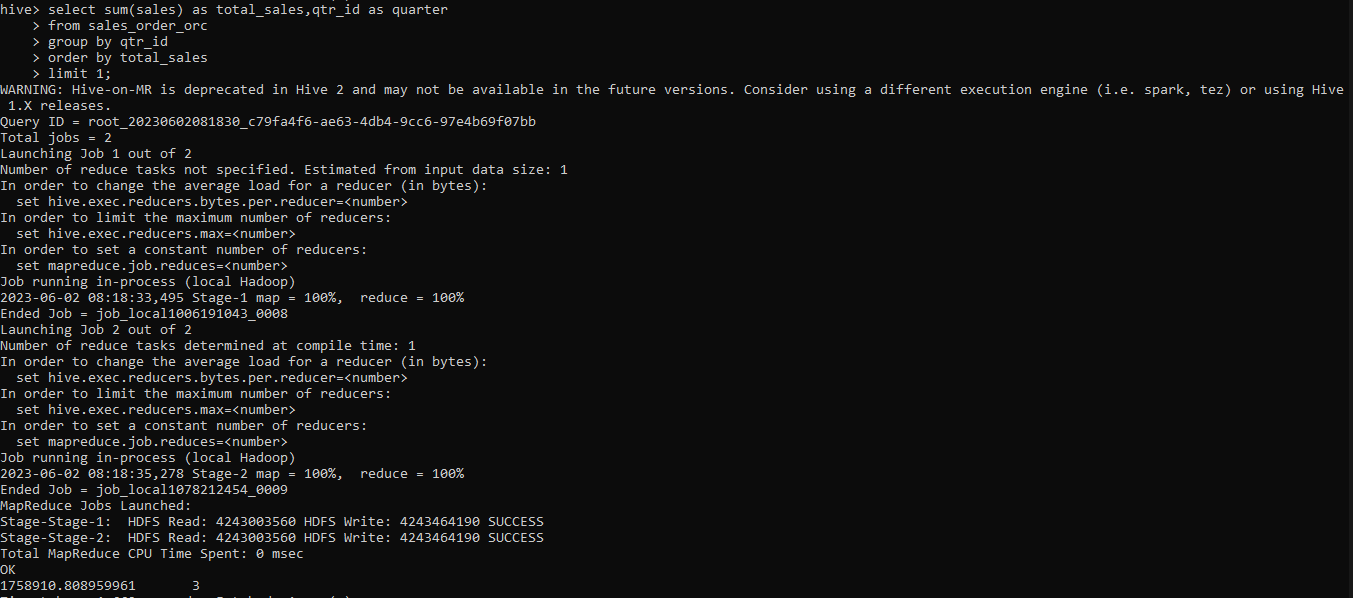
* select sum(sales) as total\_sales,qtr\_id as quarter

from sales\_order\_orc

group by qtr\_id

order by total\_sales

limit 1;

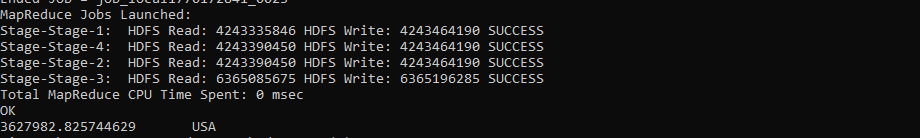


e. In which country sales was maximum and in which country sales was minimum

* (select sum(sales) as total\_sales,country from sales\_order\_orc group by country order by total\_sales limit 1)

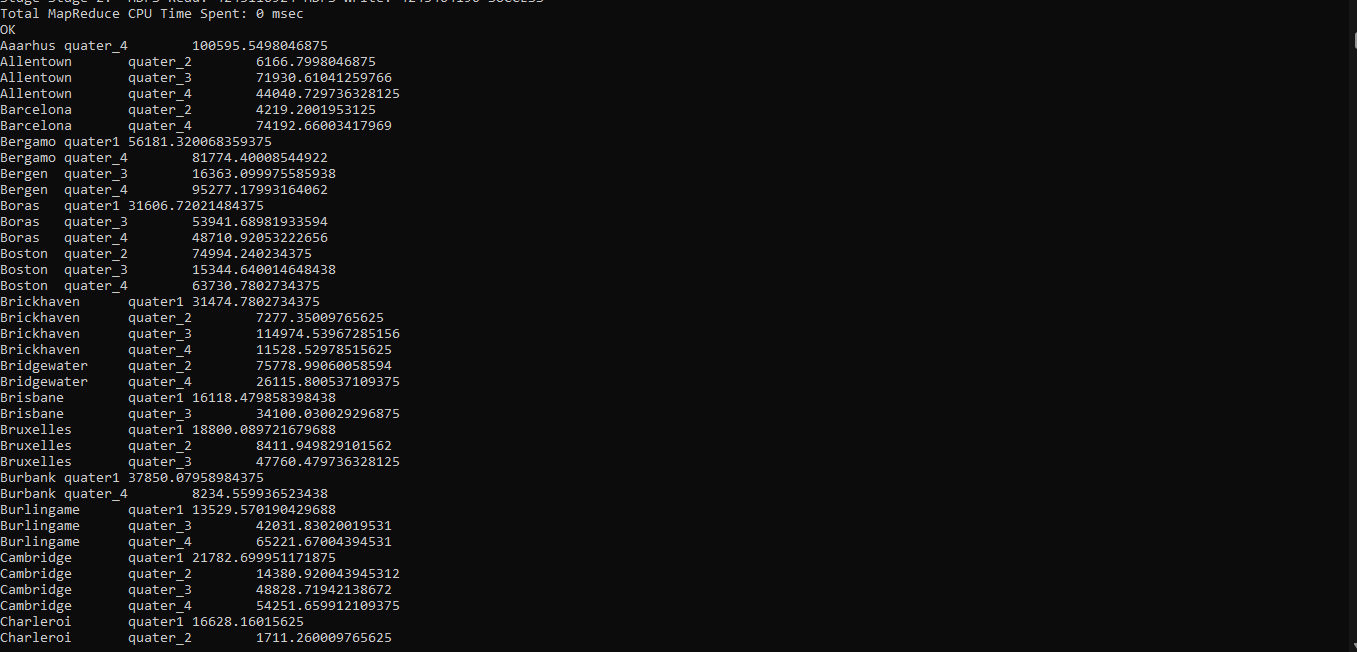
union all

select sum(sales) as total\_sales,country from sales\_order\_orc group by country order by total\_sales desc limit 1;



f. Calculate quarterly sales for each city

* select city,quarter,sum(sales) as total\_sales
* from
* (select city, case
* when month\_id >3 and month\_id < 7 then "quater\_2"
* when month\_id >6 and month\_id < 10 then 'quater\_3'
* when month\_id >9 and month\_id <13 then 'quater\_4'
* else 'quater1'
* end as quarter,
* sales
* from sales\_order\_orc
* ) t1
* group by city, quarter
* order by city, quarter;



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

* select year\_id,
* month\_id,
* total\_sales
* from
* ( select year\_id,
* month\_id,
* total\_sales,dense\_rank() over (partition by year\_id order by total\_sales desc) as ranks
* from ( select year\_id,
* month\_id,
* sum(sales) as total\_sales
* from sales\_order\_orc
* group by year\_id,
* month\_id
* ) table1
* ) table2
* where ranks = 1;

