**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**

Hadoop fs -put /config/workspace/sales\_order\_csv /uday/

**3.** **Create internal sales\_data table**

hive> create table sales\_data\_csv(

> Order\_number int, Quantity\_ordered int, Price\_each float, Orderline\_number int, Sales float,

> Status string, Qtr\_id int, Month\_id int, Year\_id int, product\_line string, msrp int, product\_code string,

> phone string,city string,state string, postal\_code string, country string,territory string,contact\_last\_name string,

> contact\_first\_name string, deal\_size string)

> row format delimited

> fields terminated by ','

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

**4. load data into internal table**

hive> load data inpath '/uday/sales\_order.csv' into table sales\_data\_csv;

**5. Create sales\_data table in ORC format**

hive> create table sales\_data\_orc(

> Order\_number int, Quantity\_ordered int, Price\_each float, Orderline\_number int, Sales float,

> Status string, Qtr\_id int, Month\_id int, Year\_id int, product\_line string, msrp int, product\_code string,

> phone string,city string,state string, postal\_code string, country string,territory string,contact\_last\_name string,

> contact\_first\_name string, deal\_size string)

> row format delimited

> fields terminated by ','

> stored as orc;

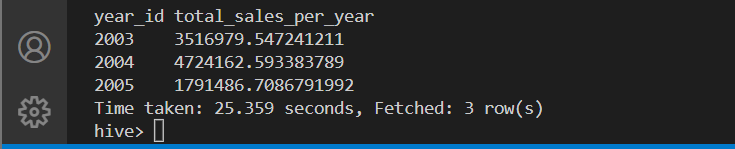
**6. load data into sales\_data\_orc from sales\_data\_csv table**

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

**PROBLEMS**

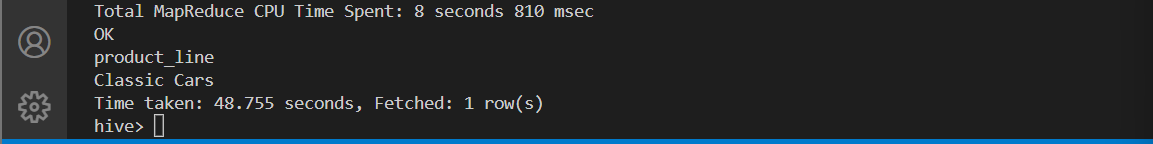
**a.Calculate total sales per year**

hive> select year\_id,sum(sales) as total\_sales\_per\_year from sales\_data\_orc group by year\_id;



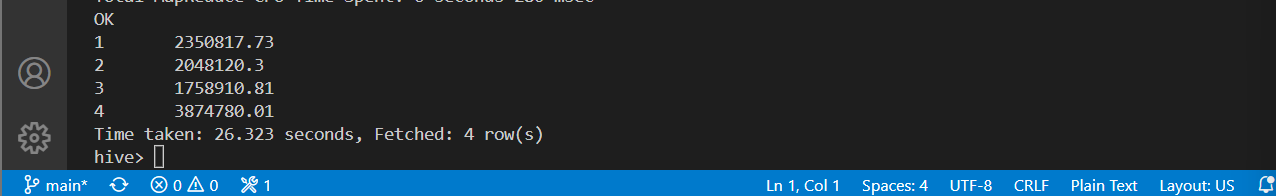
**b. Find a product for which maximum orders were placed**

hive> select product\_line from (select sum(quantity\_ordered) as high,product\_line from sales\_data\_orc group by product\_line order by high desc limit 1) as tmp;



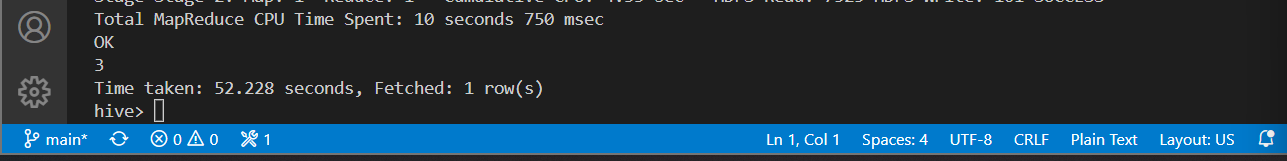
**c.** **Calculate the total sales for each quarter**

hive> select qtr\_id,round(sum(sales),2) as tot\_sales from sales\_data\_csv group by qtr\_id;



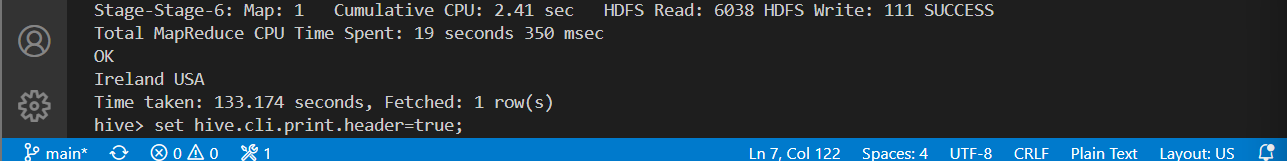
**d.** **In which quarter sales was minimum**

hive> select qtr\_id from (select qtr\_id,round(sum(sales),2) as tot\_sales from sales\_data\_csv group by qtr\_id order by tot\_sales limit 1) as tmp;



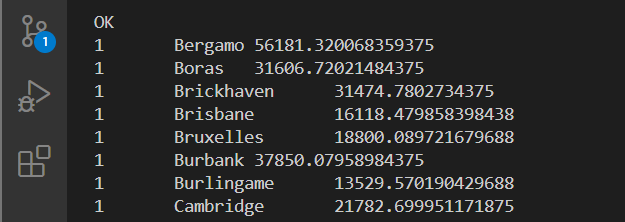
**e.Which country sales was maximum & which country sales was minimum**

hive> select mis.country as minimum\_sales ,mas.country as maximum\_sales from (select country,sum(sales) as tot\_sales from sales\_data\_orc group by country order by tot\_sales limit 1) as mis,(select country,sum(sales) as tot\_sales from sales\_data\_orc group by country order by tot\_sales desc limit 1) as mas;



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

hive> select qtr\_id,city,sum(sales) as tot\_sales from sales\_data\_orc group by city,qtr\_id ;



#ONLY FEW ARE DISPLAYED

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

hive> with tmp as (select year\_id,month\_id,row\_number()over(partition by year\_id order by sum(quantity\_ordered) desc) as row\_num from sales\_data\_orc group by year\_id,month\_id) select year\_id,month\_id from tmp where row\_num=1;

