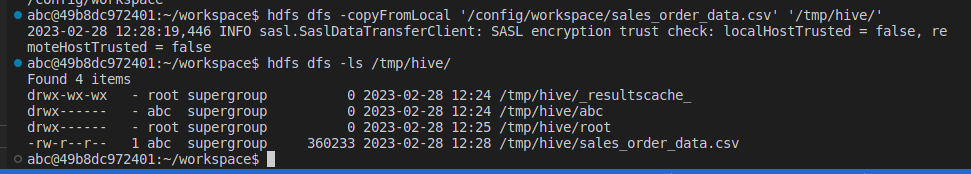
1. Download vechile sales data ->

https://github.com/shashank-mishra219/Hive-Class/blob/main/sales\_order\_data.csv

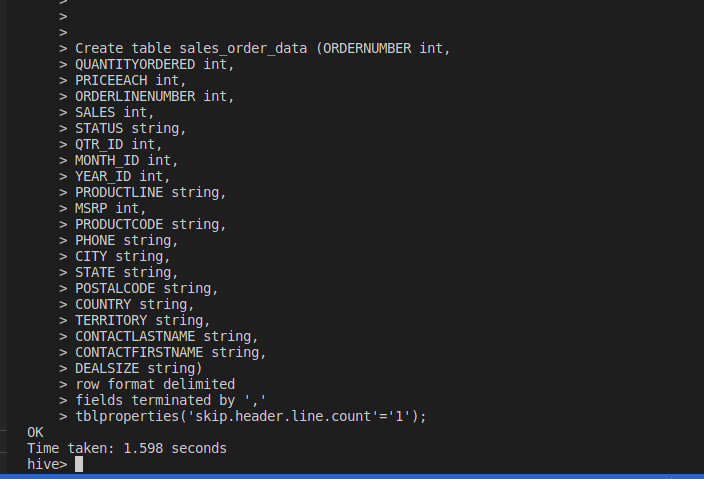
2. Store raw data into hdfs location

| hdfs dfs -copyFromLocal '/config/workspace/sales\_order\_data.csv' '/tmp/hive/' |
| --- |



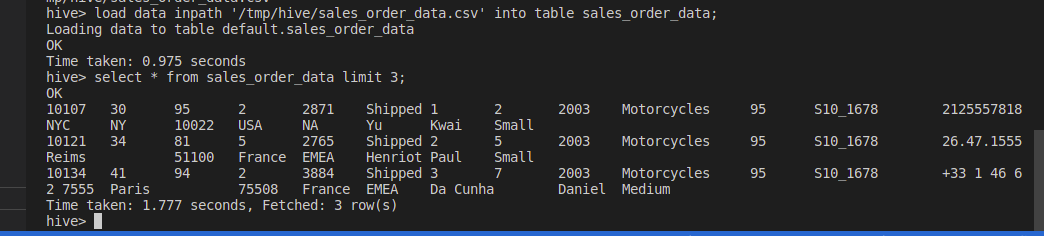
3. Create a internal hive table "sales\_order\_csv" which will store csv data sales\_order\_csv .. make sure to skip header row while creating table

| Create table sales\_order\_data (ORDERNUMBER int, QUANTITYORDERED int, PRICEEACH int, ORDERLINENUMBER int, SALES int, 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"); |
| --- |



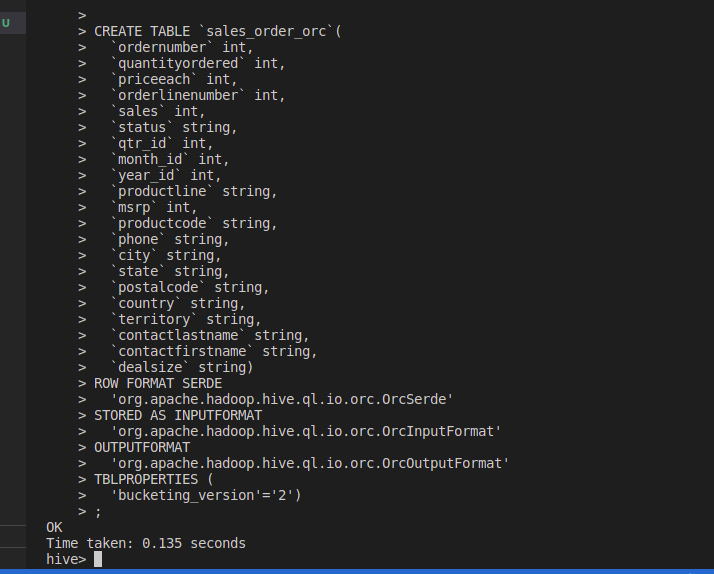
4. Load data from hdfs path into "sales\_order\_csv"

| load data inpath '/tmp/hive/sales\_order\_data.csv' into table sales\_order\_data; |
| --- |



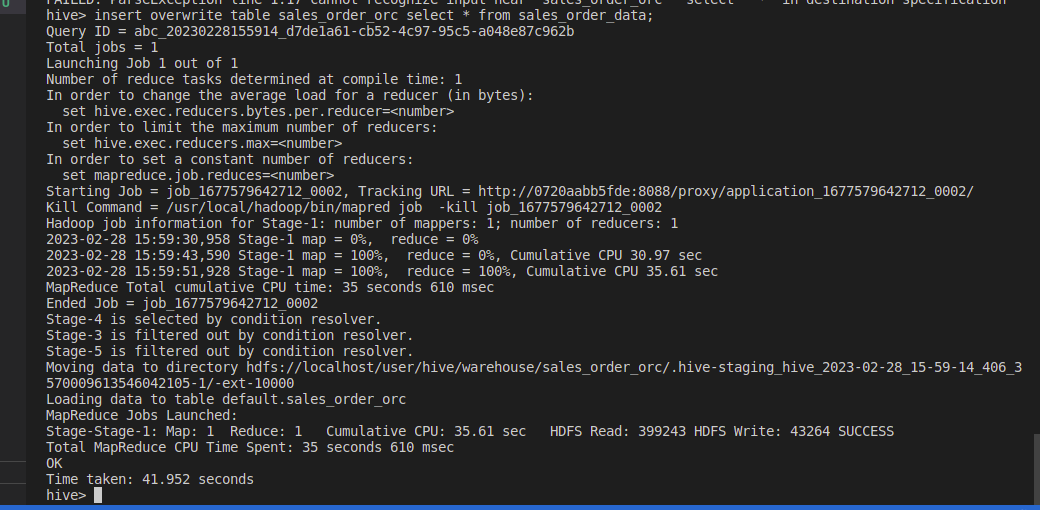
5. Create an internal hive table which will store data in ORC format "sales\_order\_orc"

| CREATE TABLE `sales\_order\_orc`(  `ordernumber` int,   `quantityordered` int,   `priceeach` int,   `orderlinenumber` int,   `sales` int,   `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 SERDE   'org.apache.hadoop.hive.ql.io.orc.OrcSerde'  STORED AS INPUTFORMAT   'org.apache.hadoop.hive.ql.io.orc.OrcInputFormat'  OUTPUTFORMAT   'org.apache.hadoop.hive.ql.io.orc.OrcOutputFormat' TBLPROPERTIES (  'bucketing\_version'='2') |
| --- |



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

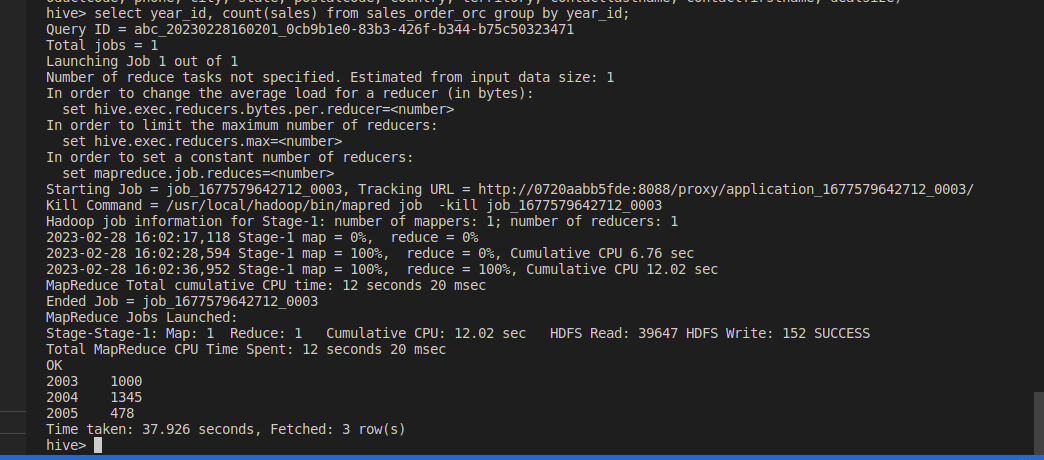
| insert overwrite table sales\_order\_orc select \* from sales\_order\_data; |
| --- |



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

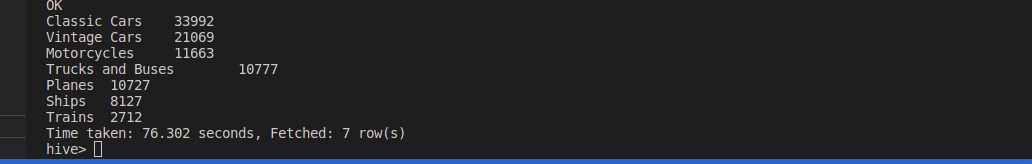
a. Calculatye total sales per year

| select year\_id, count(sales) from sales\_order\_orc group by year\_id; |
| --- |



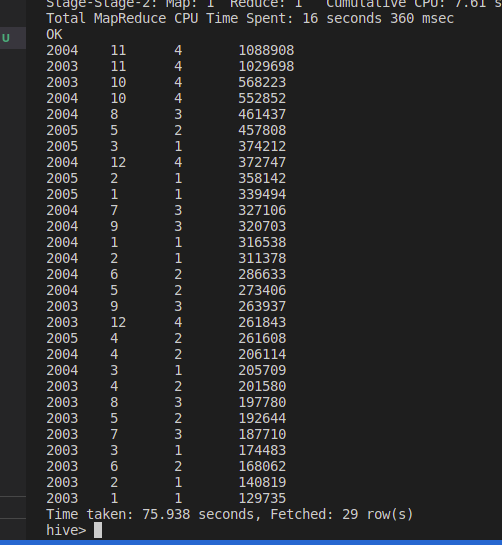
b. Find a product for which maximum orders were placed

| select PRODUCTLINE, sum(QUANTITYORDERED) from sales\_order\_orc group by PRODUCTLINE order by 2 desc; |
| --- |



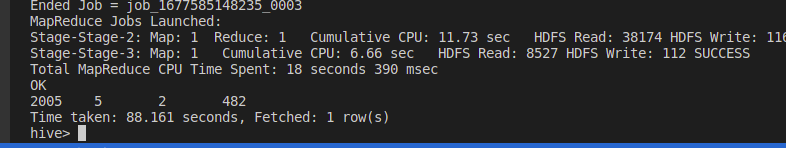
c. Calculate the total sales for each quarter

| select year\_id,month\_id, qtr\_id, sum(sales) from sales\_order\_orc group by year\_id, month\_id, qtr\_id order by 4 desc; |
| --- |



d. In which quarter sales was minimum

| Select year\_id, month\_id, qtr\_id, sales from sales\_order\_orc where sales = (select min(sales) from sales\_order\_orc); |
| --- |

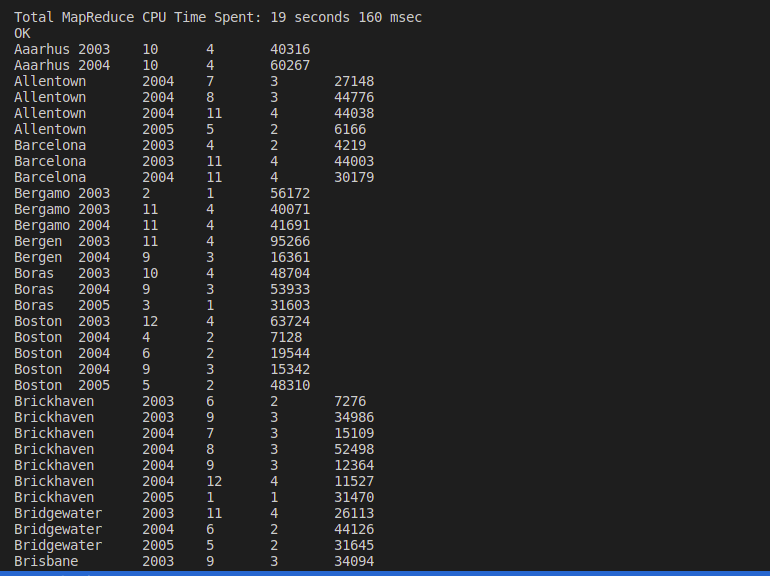


e. In which country sales was maximum and in which country sales was minimum

| Select country, sales from sales\_order\_orc where sales = (select max(sales) from sales\_order\_orc) union all Select country, sales from sales\_order\_orc where sales = (select min(sales) from sales\_order\_orc) |
| --- |

f. Calculate quartelry sales for each city

| select city,year\_id,month\_id, qtr\_id, sum(sales) from sales\_order\_orc group by city,year\_id, month\_id, qtr\_id order by 1,2,3,4 desc; |
| --- |



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

| Select year\_id, month\_id, max(month\_sale) from (Select year\_id, month\_id, sum(QUANTITYORDERED) as month\_sale from sales\_order\_orc Group by year\_id, month\_id) a; |
| --- |