Name - Shashank Yadavade
Email Id - shashankyadavade68@gmail.com
Batch - B1
Assignment - 1

Problem Statement :
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 an 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"
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"

Perform below mentioned queries on "sales_order_orc" table :
a. Calculate total sales per year
b. Find a product for which maximum orders were placed
c. Calculate the total sales for each quarter
d. In which quarter sales was minimum
e. In which country sales was maximum and in which country sales was minimum
f. Calculate quarterly sales for each city
h. Find a month for each year in which maximum number of quantities were sold

---

Solution :

Step 1:
 # Store raw data into hdfs location :
hadoop fs -put / Downloads / sales_order_data.csv hdfs_data /

Step 2:
#Create a internal hive table "sales_order_csv" which will store csv data sales_order_csv:

```
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,
contactfirstnamestring,
dealsize string
)
row format delimited
fields terminated by ','
tblproperties("skip.header.line.count"="1")
;
```

Step 3:
#Load data from hdfs path into "sales_order_csv"

hive> load data inpath ' / hdfs_data / sales_order_data.csv' into table sales_order_csv

Step 4:
#Create an internal hive table which will store data in ORC format "sales_order_orc"

hive> 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,
contactfirstnamestring,
dealsize string
)
stored as orc;

Step 4:
#Load data from "sales_order_csv" into "sales_order_orc"

hive> insert overwrite table sales_order_orc select * from sales_order_csv

---

Queries:

1 - Calculate total sales per year
Query : select year_id,sum(sales) as total_sales from sales_order_orc group by year_id ;
Output - 2003 - 3516979.547241211
2004 - 4724162.593383789
2005 - 1791486.7086791992

---

2 - Find a product for which maximum orders were placed

Query :
hive> select productline,count(ordernumber) as total_orders from sales_order_orc group by productline order by total_orders ;

Output - Trains - 77
Ships - 234
Trucks and Buses - 301
Planes - 306
Motorcycles - 331
Vintage Cars - 607
Classic Cars- 967
Ans - Classic Cars is the product for which maximum orders were placed

---

3 - Calculate the total sales for each quarter
Query:
hive> select qtr_id, sum(sales) as total_sales from sales_order_orc group by qtr_id order by qtr_id;

Ans -	**qtr_id**	**total_sales**
	1	2350817.726501465
	2	2048120.3029174805
	3	1758910.808959961
	4	3874780.010925293

---

4 - In which quarter sales was minimum?
Query:
hive> select qtr_id, sum(sales) as total_sales from sales_order_orc group by qtr_id order by total_sales;
Ans - Quarter 3 had minimum sales.

---

5 - In which country sales was maximum and in which country sales was minimum?
Query - select country, sum(sales) as total_sales from sales_order_orc group by country order by total_sales ;
Ans - Ireland has the least sales and USA has the maximum sales

Output - country	**total_sales**
Ireland	57756.43029785156
Philippines	94015.73046875
Belgium	108412.61962890625
Switzerland	117713.55859375
Japan	188167.81060791016
Austria	202062.53033447266
Sweden	210014.21020507812
Germany	220472.0897216797
Canada	224078.55993652344
Denmark	245637.15063476562
Singapore	288488.4102783203
Norway	307463.69970703125

Finland	329581.91033935547
Italy	374674.3109741211
UK	478880.45892333984
Australia	630623.0987548828
France	1110916.5217895508
Spain	1215686.9223632812
USA	3627982.825744629

---

6 - Calculate quarterly sales for each city

Query - select city,qtr_id, sum(sales) as total_sales from sales_order_orc group by city,qtr_id order by total_sales ;

---

7 - Find a month for each year in which maximum number of quantities were sold

Query - select year_id,month_id, sum(quantityordered) as total_quantities_sold from sales_order_orc group by year_id,month_id order by year_id ;

Ans -

Year	Month	Total_quantities_sold
2003	11	10179
2004	11	10678
2005	5	4357