**Name-** Subham Sahoo

**Mail id-** [sksplay7@gmail.com](mailto:sksplay7@gmail.com)

**ASSIGNMENT-1 Big data bootcamp**

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**

[cloudera@quickstart bootcamp]$ hadoop fs -mkdir /bootcamp/sales\_order\_hive

[cloudera@quickstart bootcamp]$ hadoop fs -put sales\_order\_data.csv /bootcamp/sales\_order\_hive

[cloudera@quickstart bootcamp]$ hadoop fs -ls /bootcamp/sales\_order\_hive

Found 1 items

-rw-r--r-- 1 cloudera supergroup 360233 2022-09-16 11:15 /bootcamp/sales\_order\_hive/sales\_order\_data.csv

**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\_csv\_v1

(

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,

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"**

hive>load data inpath '/bootcamp/sales\_order\_hive/' into table sales\_order\_data\_csv\_v1;

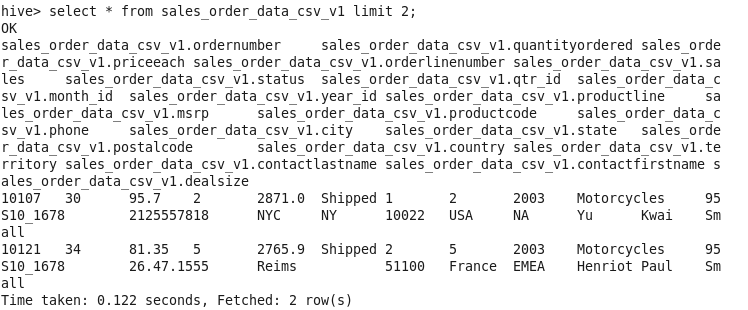
Loading data to table default.sales\_order\_data\_csv\_v1

Table default.sales\_order\_data\_csv\_v1 stats: [numFiles=1, totalSize=360233]

OK

Time taken: 0.456 seconds

hive> select \* from sales\_order\_data\_csv\_v1 limit 2;



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

create table sales\_order\_data\_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,

CONTACTFIRSTNAME string,

DEALSIZE string

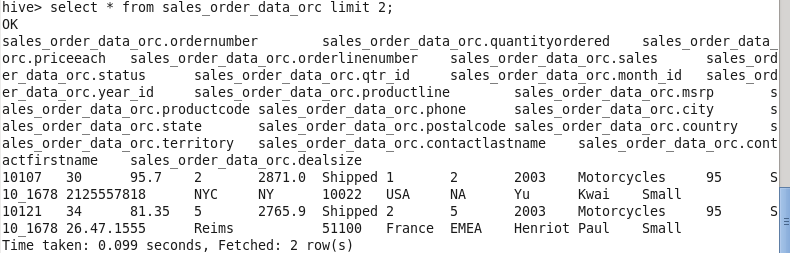
)

stored as orc;

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

hive>from sales\_order\_data\_csv\_v1 insert overwrite table sales\_order\_data\_orc select \*;

hive>select \* from sales\_order\_data\_orc limit 2;



**Perform below mentioned queries on "sales\_order\_orc" table :**

**a. Calculate total sales per year**

select year\_id, sum(sales) from sales\_order\_data\_orc group by year\_id;

year\_id \_c1

2003 3516979.547241211

2004 4724162.593383789

2005 1791486.7086791992

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

select productcode, count(\*) as orders from sales\_order\_data\_orc group by productcode order by orders desc limit 1;

productcode orders

S18\_3232 52

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

select qtr\_id, sum(sales) as total\_sales from sales\_order\_data\_orc group by qtr\_id order by qtr\_id;

qtr\_id total\_sales

1 2350817.726501465

2 2048120.3029174805

3 1758910.808959961

4 3874780.010925293

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

select qtr\_id, sum(sales) as total\_sales from sales\_order\_data\_orc group by qtr\_id order by total\_sales limit 1;

qtr\_id total\_sales

3 1758910.808959961

----- 3rd quarter

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

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

Max- USA

Min- Ireland

country total\_sales

USA 3627982.825744629

Spain 1215686.9223632812

France 1110916.5217895508

Australia 630623.0987548828

UK 478880.45892333984

Italy 374674.3109741211

Finland 329581.91033935547

Norway 307463.69970703125

Singapore 288488.4102783203

Denmark 245637.15063476562

Canada 224078.55993652344

Germany 220472.0897216797

Sweden 210014.21020507812

Austria 202062.53033447266

Japan 188167.81060791016

Switzerland 117713.55859375

Belgium 108412.61962890625

Philippines 94015.73046875

Ireland 57756.43029785156

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

select city, qtr\_id, sum(sales) as quarter\_total\_sales from sales\_order\_data\_orc group by city, qtr\_id order by city, qtr\_id;

city qtr\_id quarter\_total\_sales

Aaarhus 4 100595.5498046875

Allentown 2 6166.7998046875

Allentown 3 71930.61041259766

Allentown 4 44040.729736328125

Barcelona 2 4219.2001953125

Barcelona 4 74192.66003417969

Bergamo 1 56181.320068359375

Bergamo 4 81774.40008544922

Bergen 3 16363.099975585938

Bergen 4 95277.17993164062

Boras 1 31606.72021484375

Boras 3 53941.68981933594

Boras 4 48710.92053222656

Boston 2 74994.240234375

Boston 3 15344.640014648438

Boston 4 63730.7802734375

Brickhaven 1 31474.7802734375

Brickhaven 2 7277.35009765625

Brickhaven 3 114974.53967285156

Brickhaven 4 11528.52978515625

Bridgewater 2 75778.99060058594

Bridgewater 4 26115.800537109375

Brisbane 1 16118.479858398438

Brisbane 3 34100.030029296875

Bruxelles 1 18800.089721679688

Bruxelles 2 8411.949829101562

Bruxelles 3 47760.479736328125

Burbank 1 37850.07958984375

Burbank 4 8234.559936523438

Burlingame 1 13529.570190429688

Burlingame 3 42031.83020019531

Burlingame 4 65221.67004394531

Cambridge 1 21782.699951171875

Cambridge 2 14380.920043945312

Cambridge 3 48828.71942138672

Cambridge 4 54251.659912109375

Charleroi 1 16628.16015625

Charleroi 2 1711.260009765625

Charleroi 3 1637.199951171875

Charleroi 4 13463.480224609375

Chatswood 2 43971.429931640625

Chatswood 3 69694.40002441406

Chatswood 4 37905.14990234375

Cowes 1 26906.68017578125

Cowes 4 51334.15966796875

Dublin 1 38784.470458984375

Dublin 3 18971.959838867188

Espoo 1 51373.49072265625

Espoo 2 31018.230102539062

Espoo 3 31569.430053710938

Frankfurt 1 48698.82922363281

Frankfurt 4 36472.76025390625

Gensve 1 50432.549560546875

Gensve 3 67281.00903320312

Glen Waverly 2 14378.089965820312

Glen Waverly 3 12334.819580078125

Glen Waverly 4 37878.54992675781

Glendale 1 3987.199951171875

Glendale 2 20350.949768066406

Glendale 3 7600.1201171875

Glendale 4 34485.49987792969

Graz 1 8775.159912109375

Graz 4 43488.740234375

Helsinki 1 26422.819458007812

Helsinki 3 42744.0595703125

Helsinki 4 42083.499755859375

Kobenhavn 1 58871.110107421875

Kobenhavn 2 62091.880615234375

Kobenhavn 4 24078.610107421875

Koln 4 100306.58020019531

Las Vegas 2 33847.61975097656

Las Vegas 3 34453.84973144531

Las Vegas 4 14449.609741210938

Lille 1 20178.1298828125

Lille 4 48874.28088378906

Liverpool 2 91211.0595703125

Liverpool 4 26797.210083007812

London 1 8477.219970703125

London 2 32376.29052734375

London 4 83970.029296875

Los Angeles 1 23889.320068359375

Los Angeles 4 24159.14013671875

Lule 1 9748.999755859375

Lule 4 66005.8798828125

Lyon 1 101339.13977050781

Lyon 4 41535.11022949219

Madrid 1 357668.4899291992

Madrid 2 339588.0513305664

Madrid 3 69714.09008789062

Madrid 4 315580.80963134766

Makati City 1 55245.02014160156

Makati City 4 38770.71032714844

Manchester 1 51017.919860839844

Manchester 4 106789.88977050781

Marseille 1 2317.43994140625

Marseille 2 52481.840087890625

Marseille 4 20136.859985351562

Melbourne 1 49637.57067871094

Melbourne 2 60135.84033203125

Melbourne 4 91221.99914550781

Minato-ku 1 38191.38977050781

Minato-ku 2 26482.700256347656

Minato-ku 4 55888.65026855469

Montreal 2 58257.50012207031

Montreal 4 15947.290405273438

Munich 3 34993.92004394531

NYC 1 32647.809814453125

NYC 2 165100.33947753906

NYC 3 63027.92004394531

NYC 4 300011.6999511719

Nantes 1 59617.39978027344

Nantes 2 60344.990173339844

Nantes 3 61310.880126953125

Nantes 4 23031.589599609375

Nashua 1 12133.25

Nashua 4 119552.04949951172

New Bedford 1 48578.95935058594

New Bedford 3 45738.38952636719

New Bedford 4 113557.509765625

New Haven 2 36973.309814453125

New Haven 4 42498.760498046875

Newark 1 8722.1201171875

Newark 2 74506.06909179688

North Sydney 1 65012.41955566406

North Sydney 3 47191.76013183594

North Sydney 4 41791.949462890625

Osaka 1 50490.64013671875

Osaka 2 17114.43017578125

Oslo 3 34145.47021484375

Oslo 4 45078.759765625

Oulu 1 49055.40026855469

Oulu 2 17813.40008544922

Oulu 3 37501.580322265625

Paris 1 71494.17944335938

Paris 2 80215.4203491211

Paris 3 27798.480102539062

Paris 4 89436.60034179688

Pasadena 1 44273.359436035156

Pasadena 3 55776.119873046875

Pasadena 4 4512.47998046875

Philadelphia 1 27398.820434570312

Philadelphia 2 7287.240234375

Philadelphia 4 116503.07043457031

Reggio Emilia 2 41509.94006347656

Reggio Emilia 3 56421.650390625

Reggio Emilia 4 44669.740478515625

Reims 1 52029.07043457031

Reims 2 18971.959716796875

Reims 3 15146.31982421875

Reims 4 48895.59014892578

Salzburg 2 98104.24005126953

Salzburg 3 6693.2802734375

Salzburg 4 45001.10986328125

San Diego 1 87489.23010253906

San Francisco 1 72899.19995117188

San Francisco 4 151459.4805908203

San Jose 2 160010.27026367188

San Rafael 1 267315.2586669922

San Rafael 2 7261.75

San Rafael 3 216297.40063476562

San Rafael 4 163983.64880371094

Sevilla 4 54723.621154785156

Singapore 1 28395.18994140625

Singapore 2 92033.77014160156

Singapore 3 90250.07995605469

Singapore 4 77809.37023925781

South Brisbane 1 21730.029907226562

South Brisbane 3 10640.290161132812

South Brisbane 4 27098.800048828125

Stavern 1 54701.999755859375

Stavern 4 61897.19006347656

Strasbourg 2 80438.47985839844

Torino 3 94117.25988769531

Toulouse 1 15139.1201171875

Toulouse 3 17251.08056640625

Toulouse 4 38098.240234375

Tsawassen 2 31302.500244140625

Tsawassen 3 43332.349609375

Vancouver 4 75238.91955566406

Versailles 1 5759.419921875

Versailles 4 59074.90026855469

White Plains 4 85555.98962402344

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

select year\_id, month\_id, sum(quantityordered) as total\_quantity from sales\_order\_data\_orc group by year\_id, month\_id order by year\_id, total\_quantity desc;

Ans- 2005 (5th- May), 2004 (11th- November) and 2003 (11th- November) highest quantities sold.

year\_id month\_id total\_quantity

2003 11 10179

2003 10 5515

2003 9 2510

2003 12 2489

2003 5 2017

2003 4 1993

2003 8 1974

2003 3 1755

2003 7 1725

2003 6 1649

2003 2 1449

2003 1 1357

2004 11 10678

2004 10 5483

2004 8 4564

2004 12 3804

2004 1 3245

2004 7 3174

2004 9 3171

2004 2 3061

2004 6 2971

2004 5 2618

2004 4 2077

2004 3 1978

2005 5 4357

2005 3 3852

2005 1 3395

2005 2 3393

2005 4 2634