1. Download vehicle sales data -> <https://github.com/shashank-mishra219/Hive-Class/blob/main/sales_order_data.csv>
2. Store raw data into hdfs location

Ans : hadoop fs -put /home/dipanwita\_data/sales\_order\_data.csv /home/dipanwita\_data

hadoop fs -ls /home/dipanwita\_data

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\_agm

(

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"

load data inpath '/home/dipanwita\_data' 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\_data\_orc\_agm

(

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 ','

stored as orc;

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

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

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

1. Calculate total sales per year

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

Ans:

2003 3516979.547241211

2004 4724162.593383789

2005 1791486.7086791992

1. Find a product for which maximum orders were placed

select productline, sum(quantityordered) as total\_order from sales\_order\_data group by productline order by total\_order desc limit 1

Ans Classic Cars 33992

1. Calculate the total sales for each quarter

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

Ans:

1 2350817.726501465

2 2048120.3029174805

3 1758910.808959961

4 3874780.010925293

1. In which quarter sales was minimum

Ans 3 1758910.808959961

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

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

Maximum: Ireland 57756.43029785156

Minimum: USA 3627982.825744629

1. Calculate quartelry sales for each city

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

Minato-ku 2005 2 553.9500122070312

Stavern 2003 4 1474.6600341796875

Charleroi 2003 3 1637.199951171875

Charleroi 2003 2 1711.260009765625

Paris 2004 3 2173.60009765625

Marseille 2005 1 2317.43994140625

Singapore 2005 2 2587.5799560546875

Burlingame 2004 4 2916.199951171875

Sevilla 2004 4 3220.8800659179688

Glendale 2005 1 3987.199951171875

Singapore 2005 1 4175.60009765625

Barcelona 2003 2 4219.2001953125

Pasadena 2003 4 4512.47998046875

Glendale 2003 4 5142.14990234375

Versailles 2004 1 5759.419921875

Allentown 2005 2 6166.7998046875

Salzburg 2004 3 6693.2802734375

San Rafael 2005 2 7261.75

Brickhaven 2003 2 7277.35009765625

Philadelphia 2004 2 7287.240234375

Glendale 2003 3 7600.1201171875

Burbank 2003 4 8234.559936523438

Bruxelles 2005 2 8411.949829101562

London 2004 1 8477.219970703125

Newark 2004 1 8722.1201171875

Graz 2005 1 8775.159912109375

Frankfurt 2004 4 9214.969970703125

Lule 2003 1 9748.999755859375

South Brisbane 2003 3 10640.290161132812

Frankfurt 2003 1 11432.33984375

Brickhaven 2004 4 11528.52978515625

Nashua 2003 1 12133.25

Glen Waverly 2004 3 12334.819580078125

San Rafael 2003 1 12398.56005859375

Charleroi 2004 4 13463.480224609375

Burlingame 2005 1 13529.570190429688

London 2004 4 13739.900024414062

Paris 2005 2 14066.7998046875

Glen Waverly 2005 2 14378.089965820312

Cambridge 2004 2 14380.920043945312

Las Vegas 2004 4 14449.609741210938

Toulouse 2005 1 15139.1201171875

Reims 2003 3 15146.31982421875

Boston 2004 3 15344.640014648438

Makati City 2004 4 15928.750244140625

Montreal 2003 4 15947.290405273438

Brisbane 2004 1 16118.479858398438

Bergen 2004 3 16363.099975585938

Nantes 2003 2 16560.300048828125

Charleroi 2005 1 16628.16015625

Osaka 2004 2 17114.43017578125

Toulouse 2003 3 17251.08056640625

Oulu 2004 2 17813.40008544922

San Francisco 2003 1 18695.579833984375

Bruxelles 2004 1 18800.089721679688

Reims 2003 2 18971.959716796875

Dublin 2004 3 18971.959838867188

Marseille 2004 4 20136.859985351562

Lille 2004 1 20178.1298828125

Paris 2005 1 20321.52978515625

Glendale 2003 2 20350.949768066406

South Brisbane 2005 1 21730.029907226562

Cambridge 2004 1 21782.699951171875

Madrid 2004 3 21986.269897460938

Makati City 2003 4 22841.960083007812

Nantes 2004 4 23031.589599609375

Los Angeles 2004 1 23889.320068359375

Kobenhavn 2004 4 24078.610107421875

Los Angeles 2003 4 24159.14013671875

Singapore 2004 1 24219.58984375

Montreal 2004 2 24564.530029296875

Paris 2003 3 25624.880004882812

Minato-ku 2004 2 25928.750244140625

Kobenhavn 2005 2 26012.87060546875

Bridgewater 2003 4 26115.800537109375

Helsinki 2005 1 26422.819458007812

Boston 2004 2 26677.350219726562

Liverpool 2003 4 26797.210083007812

Cowes 2004 1 26906.68017578125

South Brisbane 2003 4 27098.800048828125

Frankfurt 2003 4 27257.790283203125

Philadelphia 2003 1 27398.820434570312

Paris 2004 2 27931.210083007812

Newark 2004 2 27987.069580078125

Chatswood 2003 3 28397.259521484375

Glendale 2004 4 29343.349975585938

Barcelona 2004 4 30183.35009765625

San Francisco 2004 4 30348.720336914062

Espoo 2004 2 31018.230102539062

Tsawassen 2004 2 31302.500244140625

Koln 2003 4 31363.1796875

Brickhaven 2005 1 31474.7802734375

Espoo 2003 3 31569.430053710938

Boras 2005 1 31606.72021484375

Bridgewater 2005 2 31648.469970703125

London 2003 2 32376.29052734375

NYC 2003 1 32647.809814453125

Montreal 2005 2 33692.97009277344

Las Vegas 2003 2 33847.61975097656

Brisbane 2003 3 34100.030029296875

Oslo 2004 3 34145.47021484375

Las Vegas 2004 3 34453.84973144531

Brickhaven 2003 3 34992.39978027344

Munich 2004 3 34993.92004394531

Strasbourg 2005 2 35680.350341796875

Kobenhavn 2004 2 36079.010009765625

Vancouver 2004 4 36576.70983886719

New Haven 2004 2 36973.309814453125

Frankfurt 2004 1 37266.48937988281

Oulu 2003 3 37501.580322265625

Burbank 2004 1 37850.07958984375

Glen Waverly 2003 4 37878.54992675781

Chatswood 2004 4 37905.14990234375

Toulouse 2003 4 38098.240234375

Minato-ku 2005 1 38191.38977050781

Paris 2003 2 38217.41046142578

Salzburg 2003 2 38629.14001464844

Vancouver 2003 4 38662.209716796875

White Plains 2003 4 38682.949462890625

Dublin 2004 1 38784.470458984375

Philadelphia 2003 4 40061.66003417969

Bergamo 2003 4 40077.71026611328

Aaarhus 2003 4 40321.60998535156

Liverpool 2005 2 40802.8095703125

Chatswood 2004 3 41297.14050292969

Reggio Emilia 2005 2 41509.94006347656

Lyon 2003 4 41535.11022949219

Bergamo 2004 4 41696.68981933594

North Sydney 2003 4 41791.949462890625

Burlingame 2003 3 42031.83020019531

Helsinki 2003 4 42083.499755859375

New Haven 2003 4 42498.760498046875

Helsinki 2004 3 42744.0595703125

Tsawassen 2004 3 43332.349609375

Graz 2003 4 43488.740234375

Singapore 2003 2 43657.47009277344

Nantes 2005 2 43784.69012451172

Chatswood 2005 2 43971.429931640625

Barcelona 2003 4 44009.30993652344

Allentown 2004 4 44040.729736328125

Bridgewater 2004 2 44130.52062988281

Singapore 2003 3 44219.36022949219

Pasadena 2005 1 44273.359436035156

Madrid 2003 1 44621.96008300781

Reggio Emilia 2003 4 44669.740478515625

Strasbourg 2004 2 44758.12951660156

Salzburg 2004 4 45001.10986328125

Oslo 2003 4 45078.759765625

New Bedford 2003 3 45738.38952636719

Singapore 2004 2 45788.72009277344

Singapore 2004 3 46030.7197265625

Newark 2005 2 46518.99951171875

White Plains 2004 4 46873.04016113281

North Sydney 2003 3 47191.76013183594

Madrid 2003 3 47727.82019042969

Bruxelles 2004 3 47760.479736328125

Boston 2005 2 48316.89001464844

New Bedford 2005 1 48578.95935058594

Boras 2003 4 48710.92053222656

Cambridge 2004 3 48828.71942138672

Lille 2003 4 48874.28088378906

Reims 2004 4 48895.59014892578

San Rafael 2004 1 48922.76989746094

Oulu 2005 1 49055.40026855469

New Bedford 2004 4 49504.379150390625

Melbourne 2004 1 49637.57067871094

San Rafael 2003 4 50360.88977050781

Liverpool 2004 2 50408.25

Gensve 2004 1 50432.549560546875

Osaka 2004 1 50490.64013671875

Manchester 2003 1 51017.919860839844

Paris 2004 1 51172.649658203125

Cowes 2004 4 51334.15966796875

Espoo 2005 1 51373.49072265625

Sevilla 2003 4 51502.74108886719

Reims 2005 1 52029.07043457031

Marseille 2003 2 52481.840087890625

Boras 2004 3 53941.68981933594

San Francisco 2005 1 54203.6201171875

Cambridge 2004 4 54251.659912109375

Stavern 2003 1 54701.999755859375

Makati City 2003 1 55245.02014160156

Nashua 2004 4 55570.59973144531

Pasadena 2003 3 55776.119873046875

Minato-ku 2004 4 55888.65026855469

Bergamo 2003 1 56181.320068359375

Reggio Emilia 2004 3 56421.650390625

Kobenhavn 2003 1 58871.110107421875

Versailles 2004 4 59074.90026855469

Salzburg 2005 2 59475.100036621094

Nantes 2005 1 59617.39978027344

Melbourne 2003 2 60135.84033203125

Aaarhus 2004 4 60273.93981933594

Stavern 2004 4 60422.530029296875

Nantes 2004 3 61310.880126953125

Burlingame 2003 4 62305.47009277344

NYC 2004 3 63027.92004394531

Boston 2003 4 63730.7802734375

Nashua 2003 4 63981.449768066406

New Bedford 2003 4 64053.130615234375

San Jose 2004 2 64600.33984375

North Sydney 2005 1 65012.41955566406

Lule 2004 4 66005.8798828125

Gensve 2004 3 67281.00903320312

Koln 2004 4 68943.40051269531

London 2003 4 70230.12927246094

NYC 2004 2 71860.779296875

Allentown 2004 3 71930.61041259766

Philadelphia 2004 4 76441.41040039062

Singapore 2003 4 77809.37023925781

Brickhaven 2004 3 79982.13989257812

San Diego 2004 1 87489.23010253906

Paris 2004 4 89436.60034179688

NYC 2003 4 89600.84008789062

Melbourne 2004 4 91221.99914550781

NYC 2003 2 93239.56018066406

San Rafael 2004 3 93928.72961425781

Torino 2004 3 94117.25988769531

Bergen 2003 4 95277.17993164062

San Jose 2005 2 95409.93041992188

Madrid 2003 2 100689.03051757812

Lyon 2004 1 101339.13977050781

Madrid 2004 1 105491.33990478516

Manchester 2004 4 106789.88977050781

Madrid 2003 4 112573.33026123047

San Rafael 2004 4 113622.75903320312

Madrid 2005 2 119242.98052978516

Madrid 2004 2 119656.04028320312

San Francisco 2003 4 121110.76025390625

San Rafael 2003 3 122368.67102050781

Madrid 2004 4 203007.4793701172

San Rafael 2005 1 205993.9287109375

Madrid 2005 1 207555.18994140625

NYC 2004 4 210410.85986328125

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

Select \* from

(Select \*,

rank() over (partition by year\_id order by quantity desc) as rk from (select year\_id, month\_id , sum(quantityordered) as quantity from sales\_order\_data group by year\_id, month\_id) as sales ) as query where rk=1;

Ans 2003 11 10179 1

2004 11 10678 1

2005 5 4357 1