**1. Download vechile sales data -> https://github.com/shashank-mishra219/Hive-Class/blob/main/sales\_order\_data.csv**

🡺Downloaded

**2. Store raw data into hdfs location**

🡺 -hdfs dfs -put 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\_csv

(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 local inpath '/sales\_order\_data.csv' into table sales\_order\_csv;

**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

)

STORED AS ORC

TBLPROPERTIES("skip.header.line.count"="1");

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

🡺 INSERT INTO TABLE sales\_order\_orc SELECT \* FROM sales\_order\_csv;

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

**a. Calculatye total sales per year**

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

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

🡺 select PRODUCTLINE,count(PRODUCTLINE) as BT from sales\_order\_orc group by PRODUCTLINE Order by BT desc limit 1;

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

🡺 select QTR\_ID,sum(sales) from sales\_order\_orc group by QTR\_ID;

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

🡺 select QTR\_ID,sum(sales)BT from sales\_order\_orc group by QTR\_ID order by BT limit1 ;

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

🡺 select COUNTRY,sum(COUNTRY)BT from sales\_order\_orc group by COUNTRY order by BT limit1 ;

select COUNTRY,sum(COUNTRY)BT from sales\_order\_orc group by COUNTRY order by BT desc limit1 ;

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

🡺select CITY,QTR\_ID,sum(sales) from sales\_order\_orc group by CITY,QTR\_ID;