Hive Assignment -1

**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**

hadoop fs -copyFromlocal "sales\_order\_data.csv" /tmp/hive\_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(

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 '/tmp/hive\_data/sales\_order\_data.csv' into table sales;

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

create table sales\_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)

row format delimited

fields terminated by ','

stored as ORC;

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

Insert overwrite table sales\_orc select \* from sales

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

* **Calculate total sales per year**

select round(sum(sales),2) as total\_sales , year\_id from sales\_orc group by year\_id;

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

select productline, sum(quantityordered) as max\_quantity from sales\_orc group by productline order by max\_quantity desc limit 1;

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

select sum(sales) as total\_Sales , qtr\_id from sales\_orc group by qtr\_id;

* **In which quarter sales was minimum**

select sum(sales) as min\_sales, qtr\_id from sales\_orc group by qtr\_id order by min\_sales desc limit 1;

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

select country, sum(sales) as maximum\_sales from sales\_orc group by country order by maximum\_sales desc limit 1;

select country, sum(sales) as minimum\_sales from sales\_orc group by country order by minimum\_sales asc limit 1;

* **Calculate quartelry sales for each city**

select count(sales) as quarterly\_sales, city,qtr\_id from sales\_orc group by city,qtr\_id order by quarterly\_sales desc;

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

select month\_id, year\_id, count(quantityordered) as qty from sales\_orc where year\_id = 2003

group by year\_id order by qty desc limit 1;

select month\_id, year\_id, count(quantityordered) as qty from sales\_orc where year\_id = 2004

group by year\_id order by qty desc limit 1;

select month\_id, year\_id, count(quantityordered) as qty from sales\_orc where year\_id = 2005

group by year\_id order by qty desc limit 1;