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

hadoop fs -mkdir /tmp/hive/hive\_class\_3\_assignment

hadoop fs -copyFromLocal /home/cloudera/data/sales\_order\_data.csv /tmp/hive/hive\_class\_3\_assignment/

**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 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//hive\_class\_3\_assignment/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 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"**

from sales\_order\_csv INSERT OVERWRITE TABLE sales\_order\_orc select \*;

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

**a. Calculate total sales per year**

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

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

select productcode,count(ordernumber) as no\_of\_orders from sales\_order\_orc

group by productcode

order by no\_of\_orders desc

limit 1;

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

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

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

select qtr\_id,sum(sales) as total\_sales from sales\_order\_orc

group by qtr\_id

order by total\_sales

limit 1;

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

**Maximum**

select country,sum(sales) as total\_sales from sales\_order\_orc

group by country

order by total\_sales desc

limit 1;

**Minimum**

select country,sum(sales) as total\_sales from sales\_order\_orc

group by country

order by total\_sales

limit 1;

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

select city,qtr\_id,sum(sales) as total\_sales

from sales\_order\_orc

group by city,qtr\_id;

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

select year\_id,month\_id,total\_quantity

from

(select

year\_id,

month\_id,

sum(quantityordered) as total\_quantity,

row\_number () over

(

partition by year\_id

order by sum(quantityordered) desc

) as rn

from sales\_order\_orc

group by year\_id,month\_id

) b

where

rn = 1;