**Assignment: 1**

**BIG DATA MASTER BATCH 1**

**1-Store raw data into hdfs location**

**Solution:**

hadoop fs -put /home/cloudera/Desktop/data/sales\_order\_data.csv /tmp/shamo;

**2-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**

**Solution:**

create table sales\_order\_csv1

> (

> order\_num int,

> quantity\_ord int,

> price\_each float,

> order\_line\_num int,

> sales float,

> status string,

> qtr\_id int,

> month\_id int,

> year\_id int,

> product\_line string,

> MSRP int,

> Product\_code string,

> phone string,

> city string,

> state string,

> Postal\_code int,

> country string,

> territory string,

> contact\_last\_name string,

> contact\_first\_name string,

> deal\_size string

> )

> row format delimited

> fields terminated by ','

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

**3- Load data from hdfs path into "sales\_order\_csv"**

**Solution:**

load data inpath '/tmp/shamo/' into table sales\_order\_csv1;

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

Solution:

create table sales\_order\_orc1

(

order\_num int,

quantity\_ord int,

price\_each float,

order\_line\_num int,

sales float,

status string,

qtr\_id int,

month\_id int,

year\_id int,

product\_line string,

MSRP int,

Product\_code string,

phone string,

city string,

state string,

Postal\_code int,

country string,

territory string,

contact\_last\_name string,

contact\_first\_name string,

deal\_size string

)

stored as orc;

**5- Load data from "sales\_order\_csv1" into "sales\_order\_orc1"**

**Solution:**

from sales\_order\_csv1 insert overwrite table sales\_order\_orc1 select \*;

**a.Calculate total sales per year**

**Solution:**

select year\_id, sum(sales) as total\_sales from sales\_order\_orc1 group by year\_id;

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

**Solution:**

select sum(quantityordered) as total, productcode from sales\_order\_orc1 group by productcode order by total desc limit 1;

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

**Solution:**

select QTR\_ID,sum(sales) from sales\_order\_orc1 group by QTR\_ID;

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

**Solution:**

select QTR\_ID,sum(sales) as Total\_Sales from sales\_order\_orc1 group by QTR\_ID order by Total\_Sales limit 1;

**E.A In which country sales was maximum**

**Solution**

select country, sum(sales) as Total\_sales from sales\_order\_orc1 group by country order by Total\_sales desc limit 1;

**E.B In which country sales was minimum**

**Solution:**

select country, sum(sales) as Total\_sales from sales\_order\_orc1 group by country order by Total\_sales limit 1;

**Calculate quartelry sales for each city**

**Solution:**

select city,QTR\_ID, sum(sales) as Total\_Sales from sales\_order\_orc1 group by QTR\_ID,city sort by QTR\_ID;