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

hadoop fs -mkdir /user/itv002776/warehouse/temp\_files

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

hadoop fs -put 'sales\_order\_data.csv' /user/itv002776/warehouse/temp\_files

A picture containing text

Description automatically generated

**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")

> ;

Text

Description automatically generated

**4. Load data from hdfs path into "sales\_order\_csv"**

load data inpath '/user/itv002776/warehouse/temp\_files/sales\_order\_data.csv' into table sales\_order\_csv;

**Graphical user interface, application, Word

Description automatically generated**

**Text

Description automatically generated**

**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;

**Table

Description automatically generated**

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

from sales\_order\_csv insert overwrite table sales\_order\_orc select \*;



Graphical user interface, text

Description automatically generated

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

1. **Calculatye total sales per year**

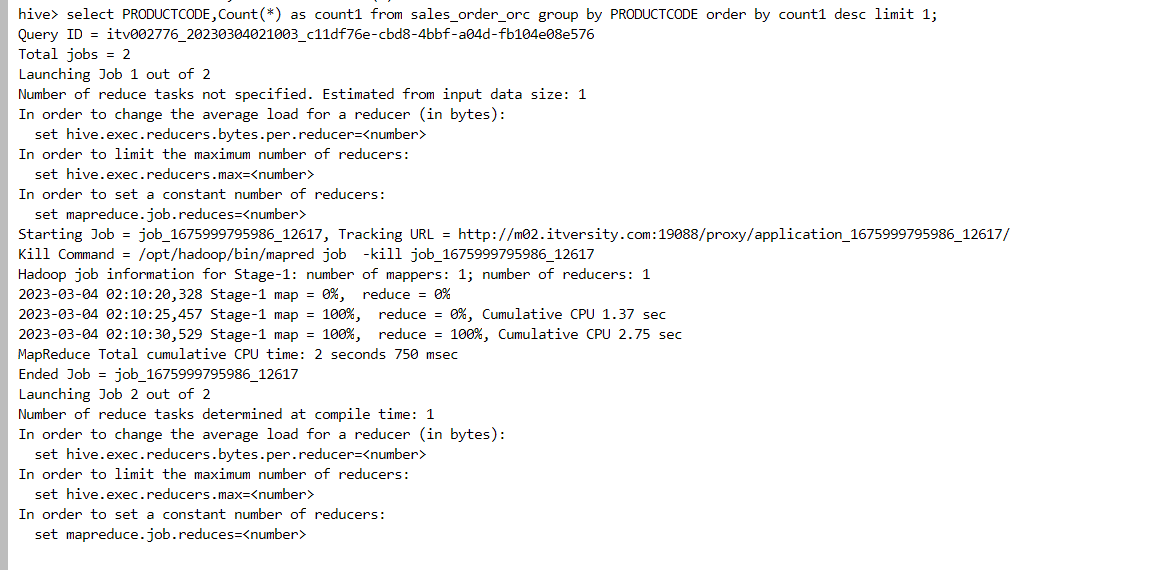
select YEAR\_ID,sum(SALES) from sales\_order\_orc group by YEAR\_ID;

Text

Description automatically generated

1. Find a product for which maximum orders were placed

select PRODUCTCODE,Count(\*) as count1 from sales\_order\_orc group by PRODUCTCODE order by count1 desc limit 1;



Graphical user interface, text, application

Description automatically generated

1. Calculate the total sales for each quarter

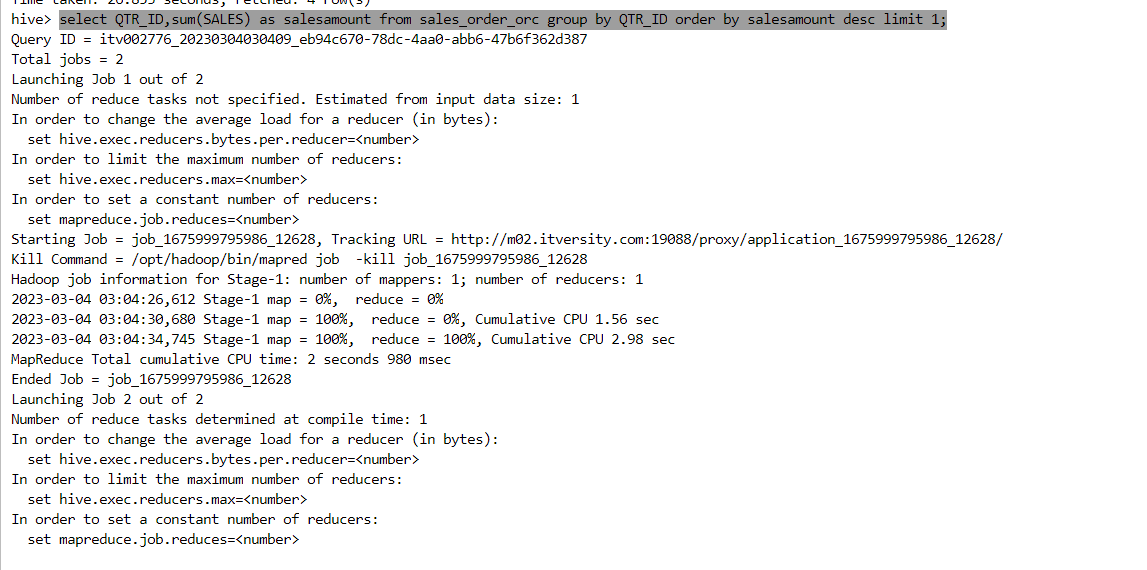
select QTR\_ID,sum(SALES) from sales\_order\_orc group by QTR\_ID;

Graphical user interface, text, application, email

Description automatically generated

1. In which quarter sales was minimum

select QTR\_ID,sum(SALES) as salesamount from sales\_order\_orc group by QTR\_ID order by salesamount desc limit 1;



Text

Description automatically generated

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

MAX: select city,sum(sales) as salesamount from sales\_order\_orc group by city order by salesamount desc limit 1;

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

MIN: select city,sum(sales) as salesamount from sales\_order\_orc group by city order by salesamount limit 1;

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

f. Calculate quartelry sales for each city.

select city,QTR\_ID,sum(SALES) as salesamount from sales\_order\_orc group by CITY,QTR\_ID order by city;

Text

Description automatically generated

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

select YEAR\_ID,max(itemsold) from (select YEAR\_ID,MONTH\_ID,count(\*) as itemsold from sales\_order\_orc group by YEAR\_ID,MONTH\_ID) temptable group by YEAR\_ID;

