****

Prashanth u

****

Assignment-1

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

2. Store raw data into hdfs location

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

4. Load data from hdfs path into "sales_order_csv"

5. Create an internal hive table which will store data in ORC format "sales_order_orc"

6. Load data from "sales_order_csv" into "sales_order_orc"

Perform below menioned queries on "sales_order_orc" table :

a. Calculatye total sales per year

select max(sales) from sales_order_csv group by YEAR_ID;

b. Find a product for which maximum orders were placed

select PRODUCTLINE from sales_order_csv where max(QUANTITYORDERED) group by PRODUCTLINE ;

c. Calculate the total sales for each quarter

SELECT sum(QTR_ID) group by YEAR_ID ;

d. In which quarter sales was minimum

SELECT min(QTR_ID) group by YEAR_ID

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

```
select max(sales),min(sales) from sales_order_csv group by COUNTRY;
```

f. Calculate quartelry sales for each city

```
SELECT sales from sales_order_csv group by CITY  ;
```

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

```
select max(MONTH_ID) from sales_order_csv group by YEAR_ID
```