1) Store raw data into hdfs location.

Solution:

To store raw data in hdfs location we can use this command.

Hdfs dfs -put sales_order_data.csv data/assignment/sales

```
Nithesh reddy@DESKTOP-CO32FOL MINGW64 /d/Nry/Big_Data_iNeuron/HiveDockerSetup (main)
$ docker exec -it namenode bash
root@8f2647c7bb8c:/# hdfs dfs -ls
ls: `.': No such file or directory
root@8f2647c7bb8c:/# hdfs dfs -ls data
ls: `data': No such file or directory
root@8f2647c7bb8c:/# hdfs dfs -ls /data
Found 3 items
drwxr-xr-x   - root supergroup          0 2023-06-20 12:30 /data/department
drwxr-xr-x   - root supergroup          0 2023-06-20 13:23 /data/employee
drwxr-xr-x   - root supergroup          0 2023-04-06 17:05 /data/openbeer
root@8f2647c7bb8c:/# hdfs dfs -mkdir -p data/assignment/sales
root@8f2647c7bb8c:/# hdfs dfs -put sales_order_data.csv /data/assignment/sales
put: `/data/assignment/sales': No such file or directory: `hdfs://namenode:9000/data/assignment/sales'
root@8f2647c7bb8c:/# hdfs dfs -put sales_order_data.csv data/assignment/sales
2023-06-20 14:10:01,633 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
root@8f2647c7bb8c:/# hdfs dfs -ls data/assignment/sales
Found 1 items
-rw-r--r--   3 root supergroup     360233 2023-06-20 14:10 data/assignment/sales/sales_order_data.csv
```

2) Create an internal hive table "sales_order_csv" which will store csv data sales_order_csv ... make sure to skip header row while creating table.

Solution:

To create an internal hive table of sales_order_csv, we have to use the following command.

create table sales_order_csv (

ordernumber int, quantityordered int, priceeach int, orderlinenumber int, sales int, 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 terminated by ‘,’;

3) Load data from hdfs path into "sales_order_csv".

Solution:

To load data from hdfs path into sales_order_csv we have use the following command:

Command: load data inpath ‘/sales_order_csv’ into table sales_order_csv

```
PROBLEMS  OUTPUT  TERMINAL  GITLENS  COMMENTS  DEBUG CONSOLE                                                    bash

hive> select * from sales_order_csv limit 20;
OK
10107   30      95.7    2       2871.0  Shipped 1       2       2003    Motorcycles     95      S10_1678        2125557818      NYC     NY      10
022     USA     NA      Yu      Kwai    Small
10121   34      81.35   5       2765.9  Shipped 2       5       2003    Motorcycles     95      S10_1678        26.47.1555      Reims           51
100     France  EMEA    Henriot Paul    Small
10134   41      94.74   2       3884.34 Shipped 3       7       2003    Motorcycles     95      S10_1678        +33 1 46 62 7555        Paris   7
5508    France  EMEA    Da Cunha        Daniel  Medium
10145   45      83.26   6       3746.7  Shipped 3       8       2003    Motorcycles     95      S10_1678        6265557265      Pasadena        CA
90003   USA     NA      Young   Julie   Medium
10159   49      100.0   14      5205.27 Shipped 4       10      2003    Motorcycles     95      S10_1678        6505551386      San Francisco   CA
        USA     NA      Brown   Julie   Medium
10168   36      96.66   1       3479.76 Shipped 4       10      2003    Motorcycles     95      S10_1678        6505556809      Burlingame      CA
94217   USA     NA      Hirano  Juri    Medium
10180   29      86.13   9       2497.77 Shipped 4       11      2003    Motorcycles     95      S10_1678        20.16.1555      Lille           59
000     France  EMEA    Rance   Martine Small
10188   48      100.0   1       5512.32 Shipped 4       11      2003    Motorcycles     95      S10_1678        +47 2267 3215   Bergen          N
5804    Norway  EMEA    Oeztan  Veysel  Medium
10201   22      98.57   2       2168.54 Shipped 4       12      2003    Motorcycles     95      S10_1678        6505555787      San Francisco   CA
        USA     NA      Murphy  Julie   Small
10211   41      100.0   14      4708.44 Shipped 1       1       2004    Motorcycles     95      S10_1678        (1) 47.55.6555  Paris           75
```

4) Create an internal hive table which will store data in ORC format "sales_order_orc"

Solution:

create table sales_order_orc (

ordernumber int, quantityordered int, priceeach int, orderlinenumber int, sales int, status string, qtr_id int, month_id int , year_id int, productline int , 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 terminated by ',' stored as orc;

5) Load data from "sales_order_csv" into "sales_order_orc

Solution: **insert overwrite table sales_order_orc select * from sales_order_cs**

6) Perform below mentioned queries on "sales_order_orc" table

a. Calculate total sales per year

```
select year_id, sum(sales) from sales_order_orc group by year_id;
```

b. Find a product for which maximum orders were placed

```
select product_line, sum(quantityordered) as maxordere from sales_order_orc group by product_line
order by maxordered desc limit 1;
```

c. Calculate the total sales for each quarter

```
select qtr_id, sum(sales) from sales_order_orc group by qtr_id;
```

d. In which quarter sales was minimum

```
select qtr_id, sum(sales) from sales_order_orc group by qtr_id order by qtr_id asc limit 1;
```

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

```
select country, sum(sales) as tot_sales from sales_order_orc group by country order by tot_sales desc
limit 1;

select country, sum(sales) as tot_sales from sales_order_orc group by country order by tot_sales asc
limit 1;
```

f. Calculate quarterly sales for each city

```
select qtr_id,city sum(sales) as qtr_sales from sales_order_orc group by qtr_id,city order by
qtr_sales;
```

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

```
select month_id as max_Q_month from (select year_id,month_id,sum(quantityordered) as cnt from
sales_order_data_orc group by year_id,month_id sort by

month_id)as cc inner join (select year_id , max(count) as cnt from (select

year_id,month_id,sum(quantityordered) as count from sales_order_data_orc group by

year_id,month_id sort by month_id) as dd group by year_id )as ee where cc.cnt =ee.cnt;
```

Nithesh Reddy M