## **Hive Assignment:**

- Download vehicle sales data -> <a href="https://github.com/shashank-mishra219/Hive-Class/blob/main/sales">https://github.com/shashank-mishra219/Hive-Class/blob/main/sales</a> order data.csv
- 2. Store raw data into hdfs location

hadoop fs -mkdir /user/cloudera/hive\_assignment

hadoop fs -put /home/cloudera/ajjoo/sales\_order\_data.csv /user/cloudera/hive\_assignment

hadoop fs -ls /user/cloudera/hive\_assignment

```
uickstart ~]$ hadoop fs -ls /user/cloudera/hive_assignment
uickstart ~]$ /home/cloudera/ajjoo/
e/cloudera/ajjoo/: is a directory
uickstart ~]$ cd /home/cloudera/ajjoo/
uickstart ~]$ cd /home/cloudera/ajjoo/
uickstart ajjoo]$ ls
csv country_wise_latest_data.csv csv_file.csv depart_data.csv hive-hcatalog-core-0.14.0.jar json_file.csv sales_data_raw.csv sales_ord
uickstart ajjoo]$ hadoop fs -put /home/cloudera/ajjoo/sales_order_data.csv /user/cloudera/hive_assignment
uickstart ajjoo]$ hadoop fs -ls /user/cloudera/hive_assignment
ms
1 cloudera cloudera 360233 2022-12-23 10:24 /user/cloudera/hive_assignment/sales_order_data.csv
```

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.

```
show databases;

create database hive_assignment;

use hive_assignment;

show tables;

set hive.cli.print.header = true;
```

```
cloudera@quickstart:~/ajjoo

hive> show databases;
OK

default
hive_class_b1
Time taken: 0.012 seconds, Fetched: 2 row(s)
hive> create database hive_assignment;
OK
Time taken: 0.192 seconds
hive> use hive_assignment;
OK
Time taken: 0.11 seconds
hive> show tables;
OK
Time taken: 0.047 seconds
hive> set hive.cli.print.header = true;
```

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 int,
COUNTRY string,
TERRITORY string,
CONTACTLASTNAME string,
CONTACTFIRSTNAME string,
DEALSIZE string
)
row format delimited
fields terminated by ','
tblproperties ("skip.header.line.count" = "1");
```

```
cloudera@quickstart:~/ajjoo
hive> 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 int,
    > COUNTRY string,
    > TERRITORY string,
    > CONTACTLASTNAME string,
    > CONTACTFIRSTNAME string,
    > DEALSIZE string
    > row format delimited
    > fields terminated by ','
    > tblproperties ("skip.header.line.count" = "1");
Time taken: 0.843 seconds
```

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

load data inpath 'hdfs:///user/cloudera/hive\_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 int,

COUNTRY string,

TERRITORY string,

CONTACTLASTNAME string,

CONTACTFIRSTNAME string,

DEALSIZE string

)

stored as ORC;
```

```
cloudera@quickstart:~/ajjoo
hive> 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 int,
    > COUNTRY string,
    > TERRITORY string,
    > CONTACTLASTNAME string,
    > CONTACTFIRSTNAME string,
    > DEALSIZE string
OK
Time taken: 0.881 seconds
hive> show tables;
OK
tab name
sales order csv
sales_order_orc
Time taken: 0.883 seconds, Fetched: 2 row(s)
hive>
```

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

```
## Conders@quickctart*

| Nive | Select sum(sales) as Total_sales, year_id from sales_order_orc group by year_id;
| Query | Do = clouders_0v22122307161e_2rDe7714-a9zf-47a0-81lb-aededacb92df
| Total_sales | Total_
```

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

Select p.productline from (Select sum(quantityordered) as maxorder,productline from sales\_order\_orc group by productline order by maxorder desc limit 1)p;

c. Calculate the total sales for each quarter

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

d. In which quarter sales was minimum?

select sum(sales) as minimum\_sales\_quarter, qtr\_id from sales\_order\_orc group by qtr\_id order by minimum\_sales\_quarter limit 1;

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

SELECT s.country, s.sales FROM(SELECT country, Sales, RANK() OVER (ORDER BY sales) rnk\_min,RANK() OVER (ORDER BY Sales DESC) rnk\_max FROM sales\_order\_orc) s WHERE rnk\_min = 1 OR rnk\_max = 1 ORDER BY Sales;

```
Addoor job information for Stage-1: number of mappers: 1: number of reducers: 1
2022-12-20 06:58:11,498 Stage-1 map = 004, reduce = 08, Comulative CFU 2.99 sec
2022-12-20 06:58:10,198 Stage-1 map = 1004, reduce = 08, Comulative CFU 2.99 sec
2022-12-20 06:58:10,198 Stage-1 map = 1004, reduce = 08, Comulative CFU 1.8 sec
2022-12-20 06:58:10,198 Stage-1 map = 1004, reduce = 08, Comulative CFU 1.8 sec
2022-12-20 06:58:10,198 Stage-1 map = 1004, reduce = 08, Comulative CFU 1.8 sec
2022-12-20 06:58:10,198 Stage-1 map = 1004, reduce = 08, Comulative CFU 1.8 sec
2022-12-20 06:58:10,198 Stage-1 map = 1004, reducer(in bytes):
2022-12-20 06:58:10,198 Stage-1 map = 1004, reducer(in bytes):
2022-12-20 06:58:10,198 Stage-2 map = 004, reducer(in bytes):
2022-12-20 06:59:00,598 Stage-2 map = 004, reducer(in byt
```

f. Calculate quartelry sales for each city.Select sum(sales) as quarterly\_sales, qtr\_id, city from sales\_order\_orc group by qtr\_id, city;

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

select month\_id,year\_id,QUANTITYORDERED from (select
month\_id,year\_id,QUANTITYORDERED ,dense\_rank() over(partition by year\_id order
by QUANTITYORDERED desc) as rnks from sales\_order\_csv)s where s.rnks =1;