## hive\_class\_3\_assignment

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

Sol: Downloaded and transferred file into cloudera VM.

[cloudera@quickstart ~]$ ls /tmp/hive\_class/sales\_order\*

/tmp/hive\_class/sales\_order\_data.csv

2. Store raw data into hdfs location

Sol: Move data from local file system to HDFS.

[cloudera@quickstart ~]$ hdfs dfs -put /tmp/hive\_class/sales\_order\_data.csv /tmp/hive\_class/

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

Sol:

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

fields terminated by ','

tblproperties("skip.header.line.count"="1");

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

Sol: load data inpath '/tmp/hive\_class/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"

Sol:

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 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 ','

stored as ORC;

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

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

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

1. Calculate total sales per year

Sol: year\_id total\_sales

1 2003 3516514

2 2004 4723531

3 2005 1791264

1. Find a product for which maximum orders were placed

Sol: SELECT productcode, count(\*) as cnt from sales\_order\_orc group by productcode ORDER BY cnt DESC limit 1;

productcode cnt

S18\_3232 52

1. Calculate the total sales for each quarter

Sol: SELECT qtr\_id, sum(sales) as totalsales from sales\_order\_orc GROUP BY qtr\_id;

qtr\_id totalsales

1 2350510

2 2047855

3 1758673

4 3874271

d. In which quarter sales was minimum

Sol:SELECT qtr\_id, sum(sales) as totalsales from sales\_order\_orc GROUP BY qtr\_id ORDER BY totalsales limit 1;

qtr\_id totalsales

3 1758673

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

Sol:

SELECT country, sum(sales) as totalsales from sales\_order\_orc GROUP BY country ORDER BY totalsales limit 1;

Minimum: Ireland 57749

SELECT country, sum(sales) as totalsales from sales\_order\_orc GROUP BY country ORDER BY totalsales desc limit 1;

Maximum: USA 3627511

f. Calculate quartelry sales for each city

Sol:select city, qtr\_id, sum(sales) as salesforeachquarter from sales\_order\_orc group by qtr\_id,city order by city;

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

Sol: select month\_id, year\_id, max(quantityordered) as quantity from sales\_order\_orc group by year\_id, month\_id order by year\_id;

Note: I am not able perform last query properly, I am getting all records of each month and year