**Hive Assignment-1**

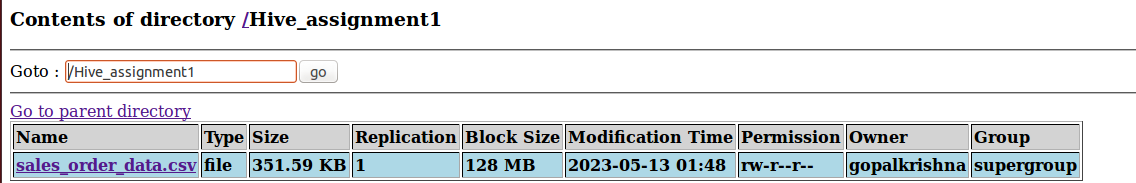
1. **Download vechile sales data ->** <https://github.com/shashank-mishra219/Hive-Class/blob/main/sales_order_data.csv>
2. **Store raw data into hdfs location.**

gopalkrishna@ubuntu:~/KRP$ vi sales\_order\_data.csv

(Note: Header removed here)

gopalkrishna@ubuntu:~/KRP$ hdfs dfs -mkdir /Hive\_assignment1/

gopalkrishna@ubuntu:~/KRP$ hdfs dfs -put /home/gopalkrishna/KRP/sales\_order\_data.csv /Hive\_assignment1



1. **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**.

hive> create table sales\_order\_csv(Ordernumber int,Quantity\_ordered int,Price\_each float,Order\_line\_number int,Sales float,Status string,Qtr\_id int,Month\_id int,Year\_id int,Productline string,MSRP int,Product\_code string,Phone string,City string,State string,Postal\_code string,Country string,Territory string,Contact\_last\_name string,Contact\_first\_name string,Deal\_size string)

> row format delimited

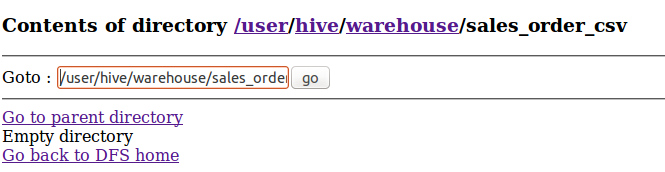
> fields terminated by ','

> lines terminated by '\n'

> stored as textfile;

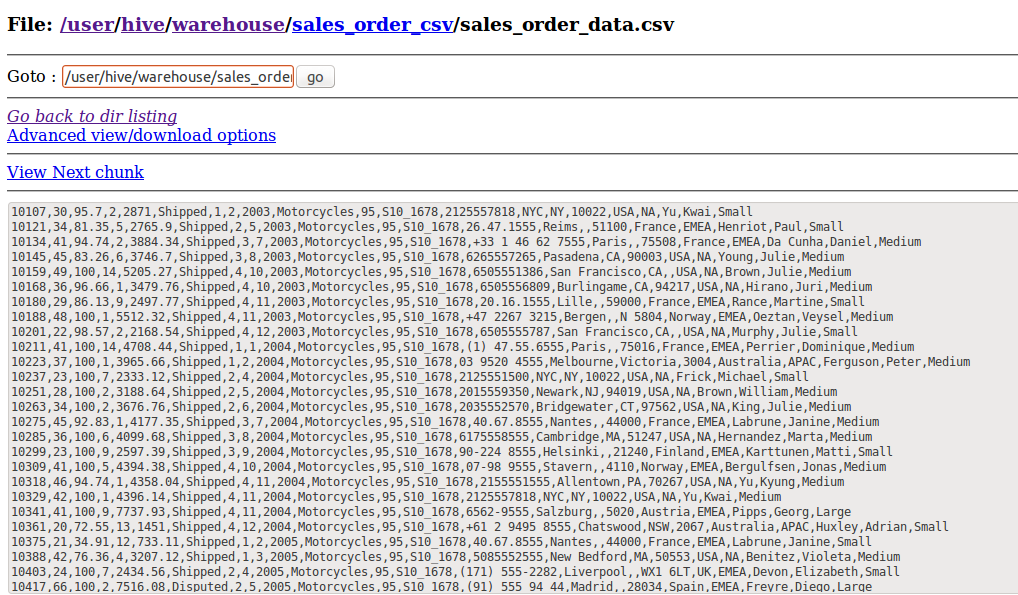
OK

Time taken: 0.888 seconds



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

hive> load data inpath '/Hive\_assignment1/sales\_order\_data.csv' into table sales\_order\_csv;



1. **Create an internal hive table which will store data in ORC format "sales\_order\_orc".**

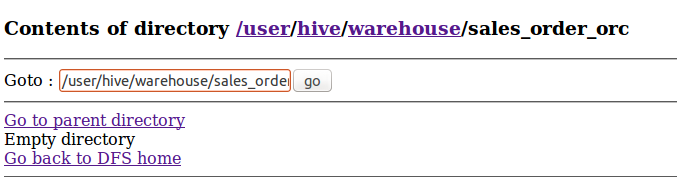
hive> create table sales\_order\_orc(Ordernumber int,Quantity\_ordered int,Price\_each float,Order\_line\_number int,Sales float,Status string,Qtr\_id int,Month\_id int,Year\_id int,Productline string,MSRP int,Product\_code string,Phone string,City string,State string,Postal\_code string,Country string,Territory string,Contact\_last\_name string,Contact\_first\_name string,Deal\_size string)

> row format delimited

> fields terminated by ','

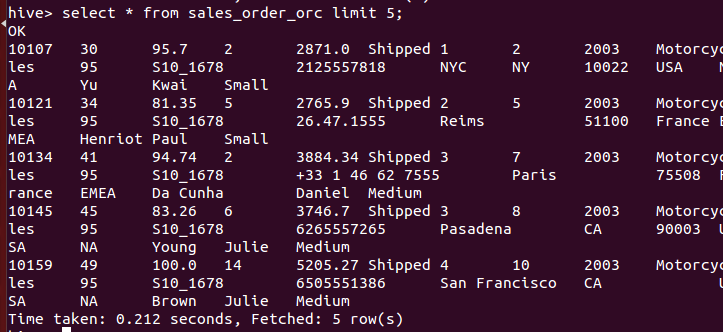
> lines terminated by '\n'

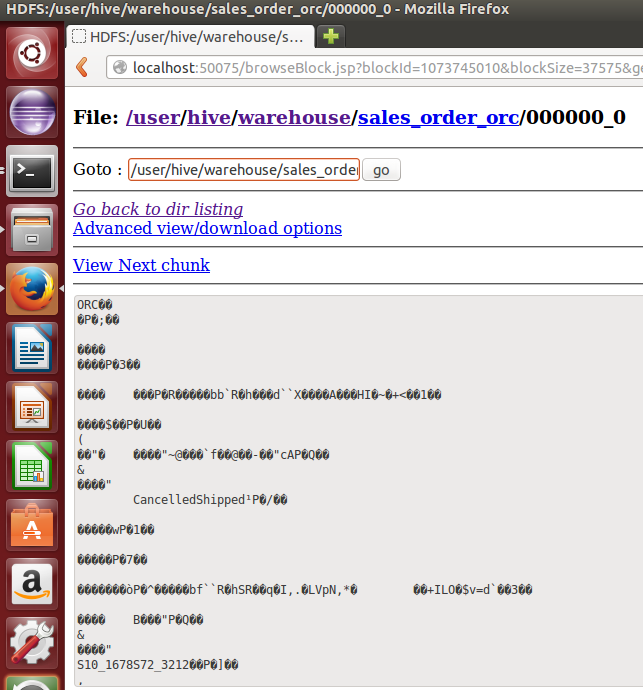
> stored as orc;



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

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

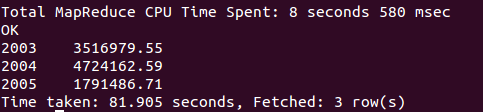




**🡺 Perform below menioned queries on "sales\_order\_orc" table** :

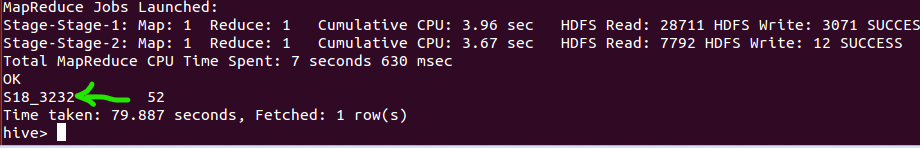
1. **Calculatye total sales per year:**

hive> select Year\_id, Round(sum(Sales), 2) as Total\_sales\_per\_year from sales\_order\_orc group by Year\_id order by Year\_id;

****

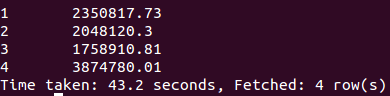
1. **Find a product for which maximum orders were placed.**

hive> select Product\_code,count(Ordernumber) as Max\_ordered\_product from sales\_order\_orc group by Product\_code order by Max\_ordered\_product desc limit 1;



1. **Calculate the total sales for each quarter.**

hive> select Qtr\_id,round(sum(Sales),2) as sales\_per\_qtr from sales\_order\_orc group by Qtr\_id;



1. **In which quarter sales was minimum.**

hive> select Qtr\_id,round(sum(Sales),2) as sales\_per\_qtr from sales\_order\_orc group by Qtr\_id order by sales\_per\_qtr limit 1;

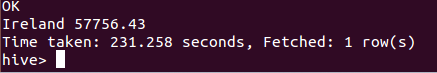


1. **In which country sales was maximum and in which country sales was minimum.**

hive> select Country,round(sum(Sales),2) as sales\_per\_country from sales\_order\_orc group by Country order by sales\_per\_country desc limit 1;

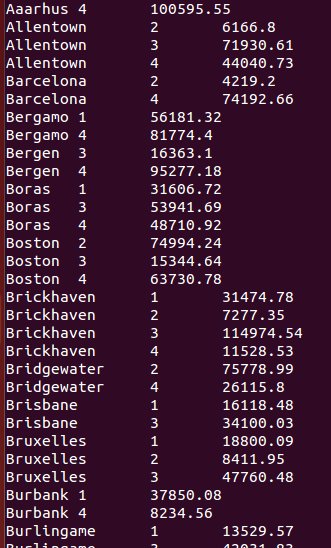


hive> select Country,round(sum(Sales),2) as sales\_per\_country from sales\_order\_orc group by Country order by sales\_per\_country limit 1;

****

1. **Calculate quartelry sales for each city.**

hive> select City,Qtr\_id,round(sum(Sales),2) from sales\_order\_orc group by City,Qtr\_id order by City,Qtr\_id;



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

hive> with abc as (select year\_id, month\_id, sum(quantity\_ordered) as qty from sales\_order\_orc group by year\_id, month\_id),xyz as (select year\_id, month\_id, qty,row\_number() over(partition by year\_id order by year\_id, qty desc) ranking from abc) select year\_id, month\_id, qty as Qty\_sold from xyz where ranking = 1;

