HIVE QUERIES

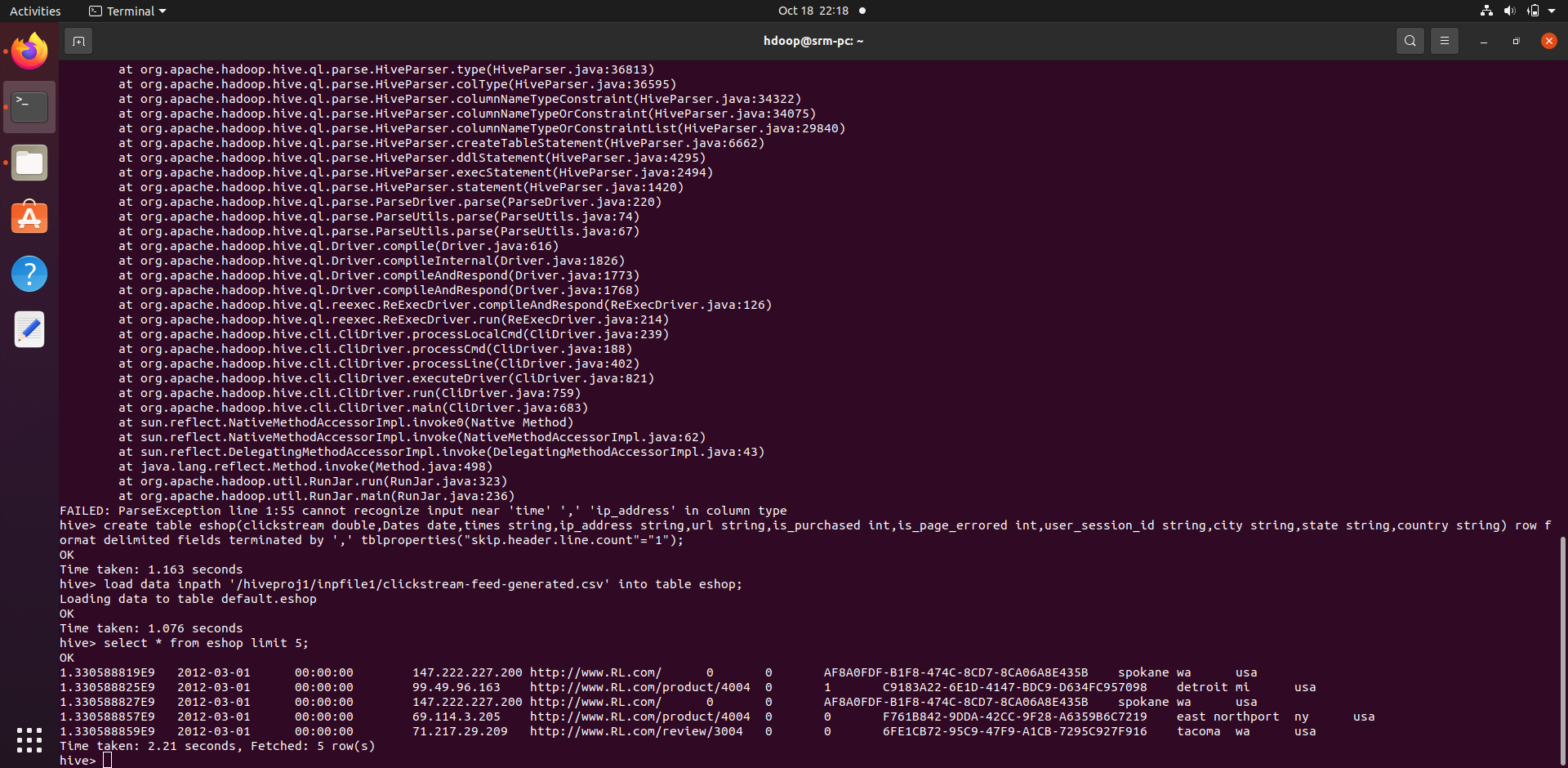
**Hive Queries:**

Insertion of dataset:

create table eshop(clickstream double,datetime timestamp,ip\_address string,url string,is\_purchased int,is\_page\_errored int,user\_session\_id string,city string,state string,country string) row format delimited fields terminated by ',' tblproperties("skip.header.line.count"="1");

output:

This query is used to create table in a database.



Working of hive:

Select \* from eshop limit 10;

Output:

This query is used to print or view 10 instances from loaded dataset.

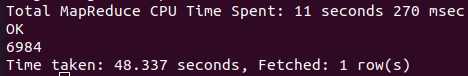


**Queries**

1. Query : Select count(is\_page\_errored) from eshop where is\_page\_errored==1;

Output:

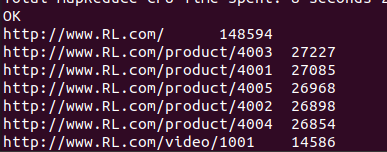
This query gets the total count of the web pages which crashed or errored during a customer purchase



1. Query : Select url,count(\*) as max\_count from eshop group by url order by count(\*) desc limit 10;

Output:

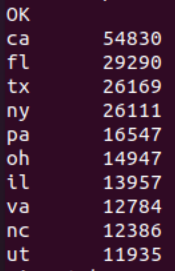
This query is used to get the results of the most viewed web page among all the available web pages



1. Query : Select state,count(\*) as max\_count from eshop group by state order by count(\*) desc limit 10;

Output:

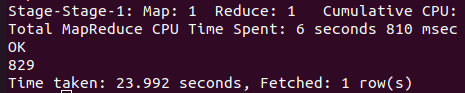
This query gets the count of the customers from various cities in the country



1. Query : Select count(DISTINCT(IP\_address)) from eshop where Dates='2012-03-01';

Output:

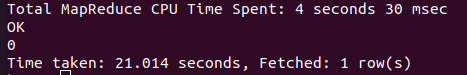
This query is used to get the number of customers who visited our web page on the day dated ‘2012-03-01’



1. Query : Select count(DISTINCT(IP\_address)) from eshop where day('2012-04-15')-3=day('2012-04-10');

Output:

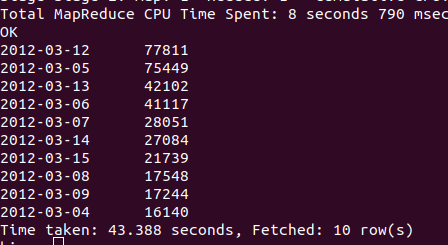
This query is used to check whether any new customer have visited our web page in the past three days



1. Query : Select Dates,count(IP\_address) as max\_count from eshop group by Dates order by count(\*) desc limit 10;

Output:

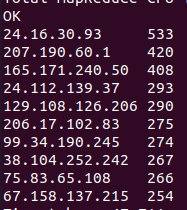
This query gets the dates where the interaction has been the highest printing them from highest to the lowest



1. Query : Select ip\_address,count(is\_purchased) as max\_count from eshop group by ip\_address order by count(\*) desc limit 10;

Output:

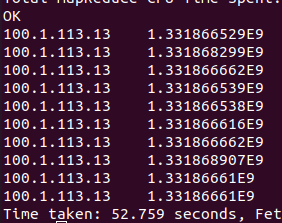
This query is used to get the IP address the user who has purchased the most from our website



1. Select IP\_address , clickstream\_id from eshop CLUSTER BY IP\_address limit 10;

Output:

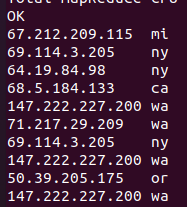
This query has “clustered by” clause is used to divide the table into buckets.



1. Select IP\_address , state from eshop DISTRIBUTE BY state limit 10;

Output:

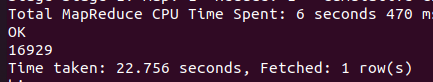
This query has DISTRIBUTE BY clause is used to distribute the input rows among reducers.



10) Select sum(is\_purchased) from eshop;

Output:

This query gets the quantity of total purchases that occured



11) select sum(is\_page\_errored) from eshop;

Output:

This query gets the the total number of times an error occurred while viewing or purchasing from our website

