CREATE EXTERNAL TABLE CCA (Product STRING,Sub\_Product STRING,Company STRING,State STRING ,Zipcode INT, Submitted\_via STRING, Company\_responsetoconsumer STRING,Complaint\_ID INT)

ROW FORMAT DELIMITED fields terminated by ‘,’

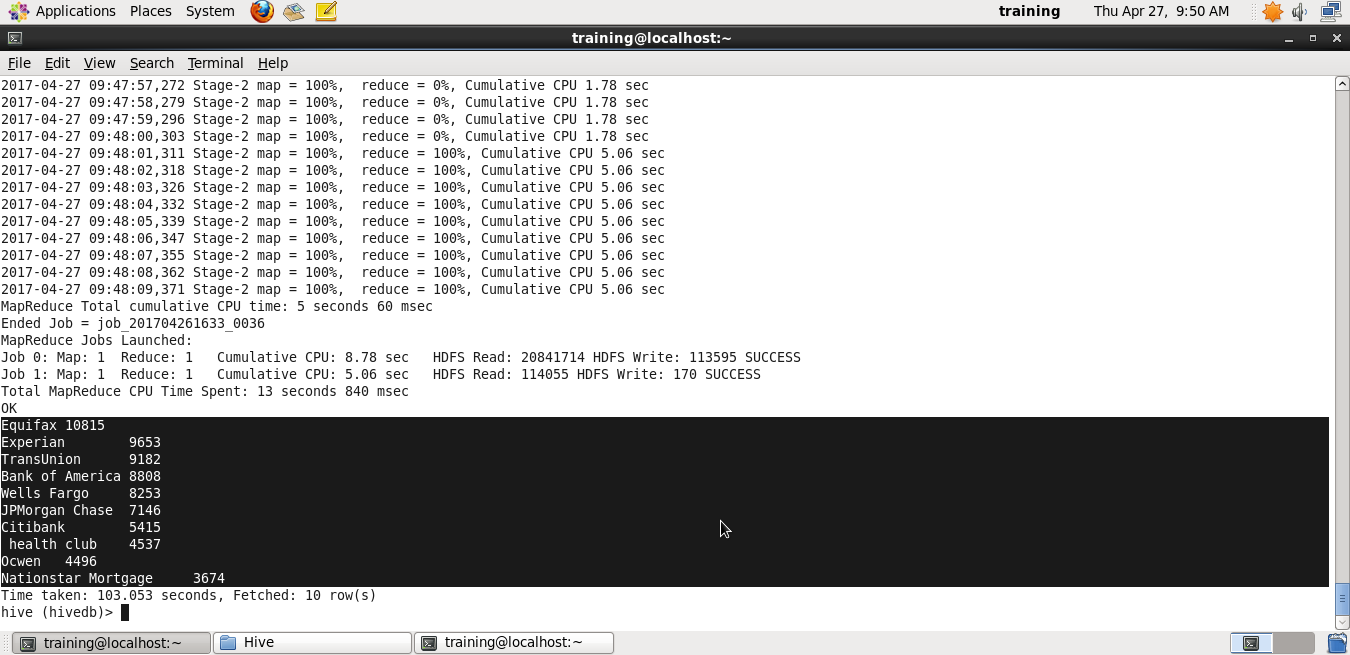
location '/home/training/';

**Load data into CCA**

Problem Statement: Analyze the data in Hadoop Eco-system to:

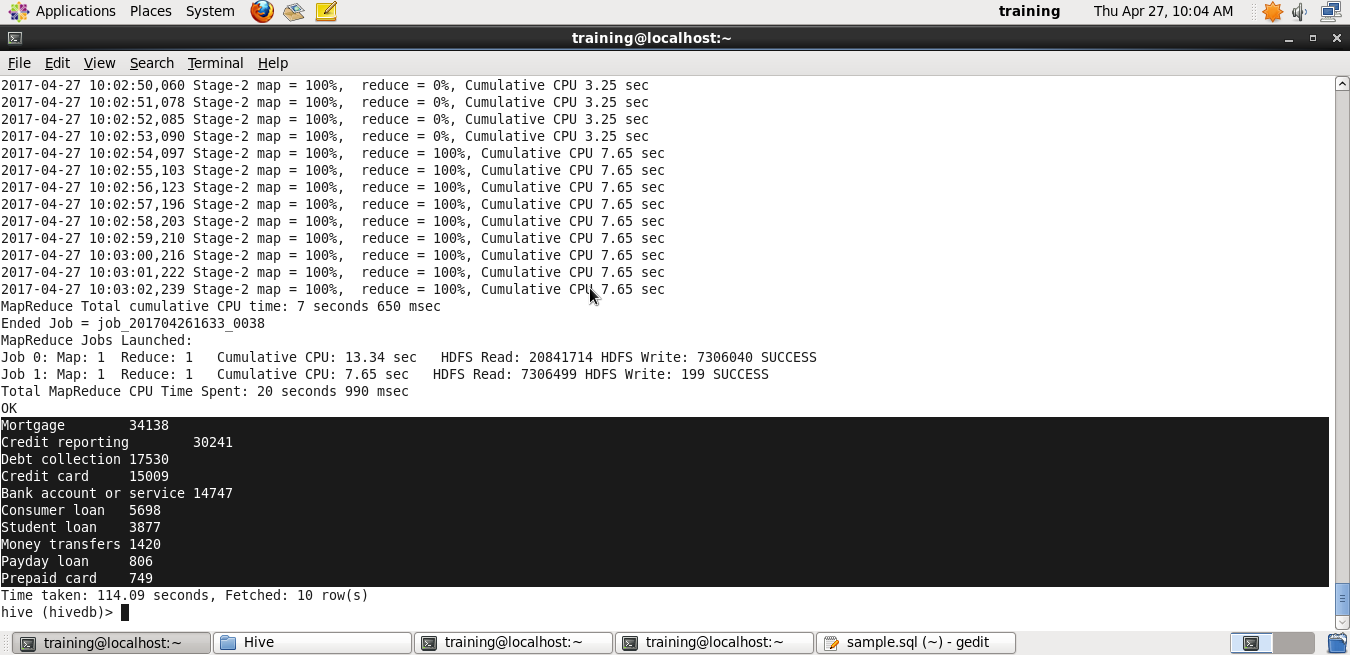
1. **Get the number of complaints filed for each company.**

Select Company, count(Complaint\_ID) as cnt from cca group by Company order by cnt desc limit 10;



**2. Get the number of complaints filed under each product.**

Select Product, count(Complaint\_ID) as cnt from cca group by Product order by cnt desc limit 10;



**3. Get the total number of complaints filed from a particular location**

Select Product, count(Complaint\_ID) as cnt from cca group by Product order by cnt desc limit 10;

To run these commands as Hive script:

gedit sample.sql

and then enter the commands

Select Company, count(Complaint\_ID) as cnt from cca group by Company order by cnt desc limit 10;

Select Product, count(Complaint\_ID) as cnt from cca group by Product order by cnt desc limit 10;

Select Product, count(Complaint\_ID) as cnt from cca group by Product order by cnt desc limit 10;

**hive -f '/home/training/Desktop/Hive/sample.sql';**