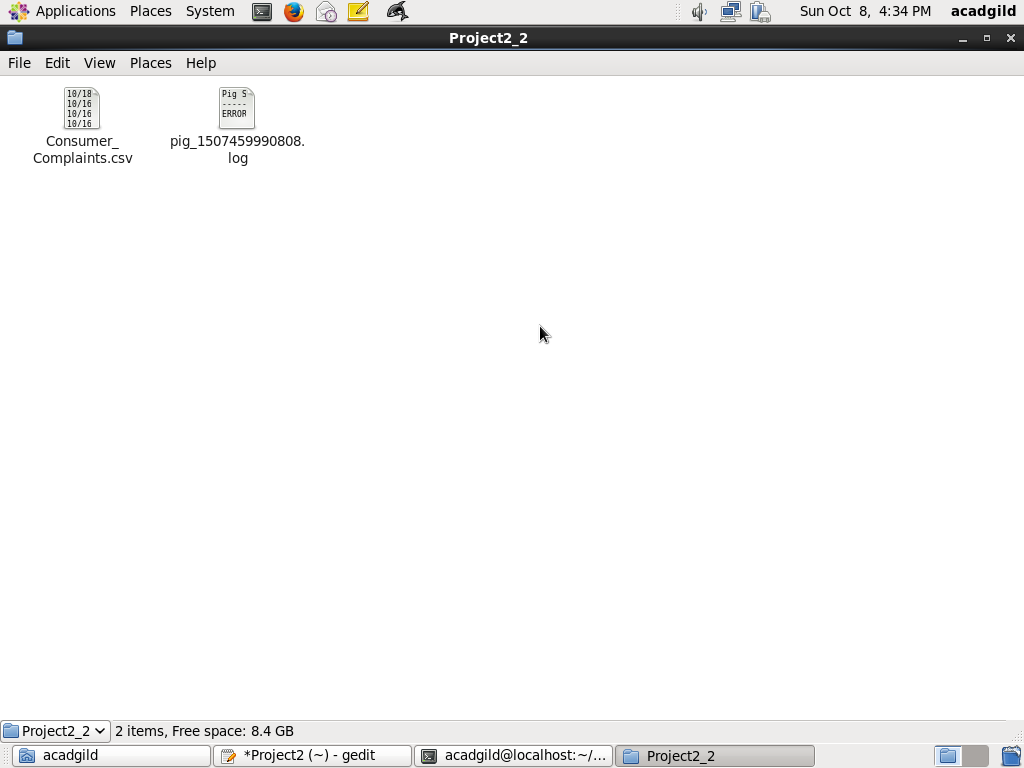
**Task 1:**

**To copy the dataset into HDFS using Flume**

flume-ng agent --conf-file flume.conf --name agent1



**Task 2**

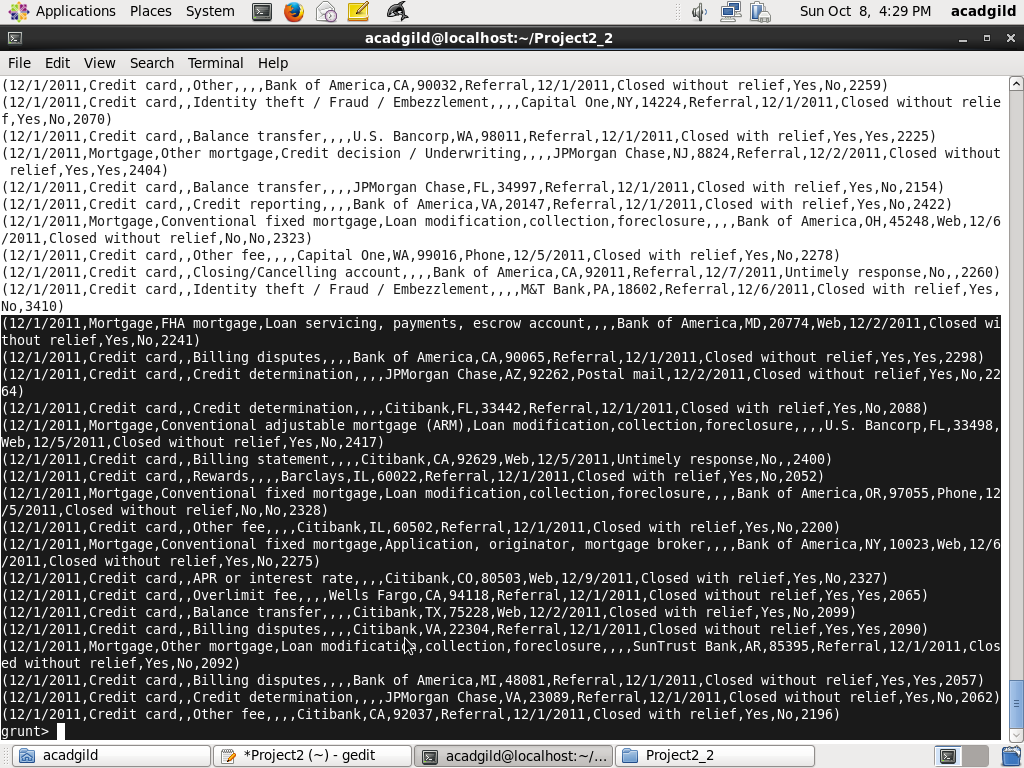
**Write a pig script to find no of complaints which got timely response and copy the results into mysql using sqoop export**

REGISTER ‘home/acadgild/pig-0.14.0/lib/piggybank.jar’;

DEFINE CSVLoader org.apache.pig.piggybank.storage.CSVLoader();

A= load ‘user/acadgild/Project2/Consumer\_Complaints.csv’ using CSVLoader(‘,’);

Dump A;



B = Foreach A generate $15 as id, $13 as timely;

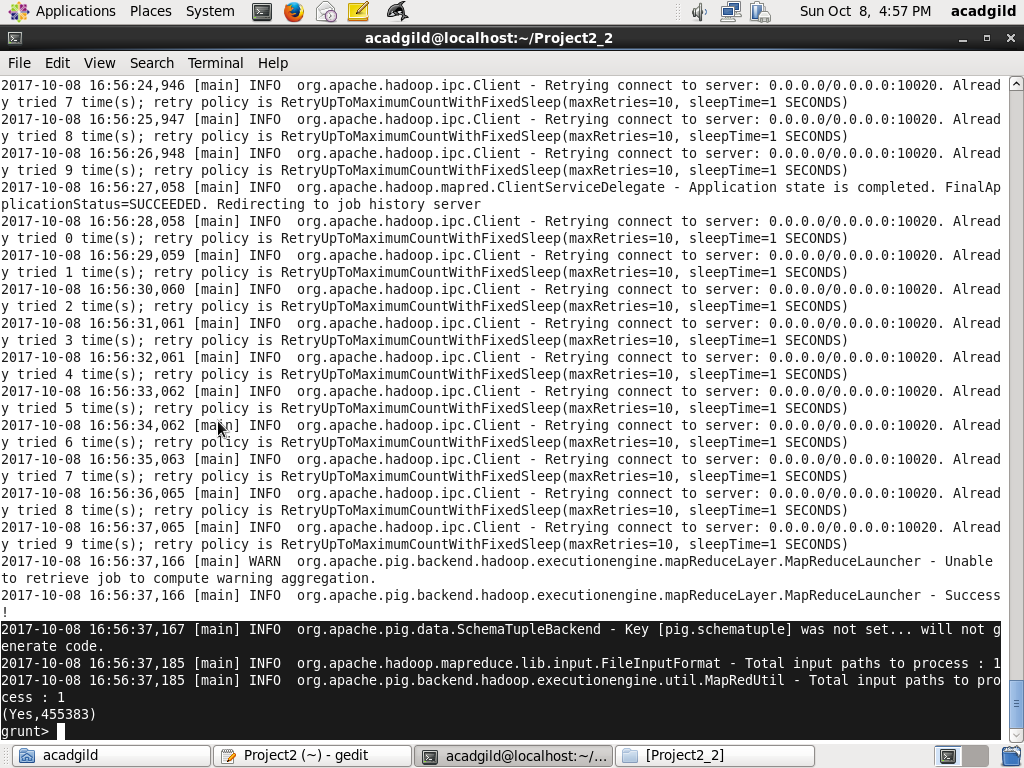
C= Filter B by timely=’Yes’;

D = Foreach C generate id,timely;

E = group D by timely;

F = Foreach E generate group as timely , COUNT(D.id) as numofcomplaints;

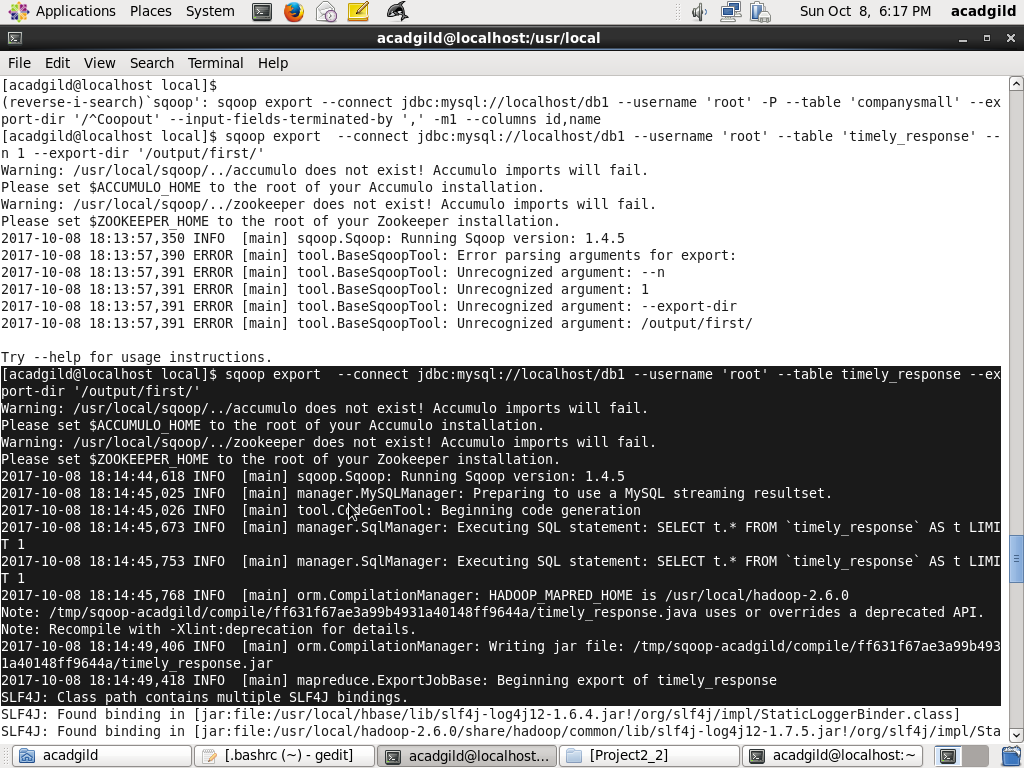
Dump F;



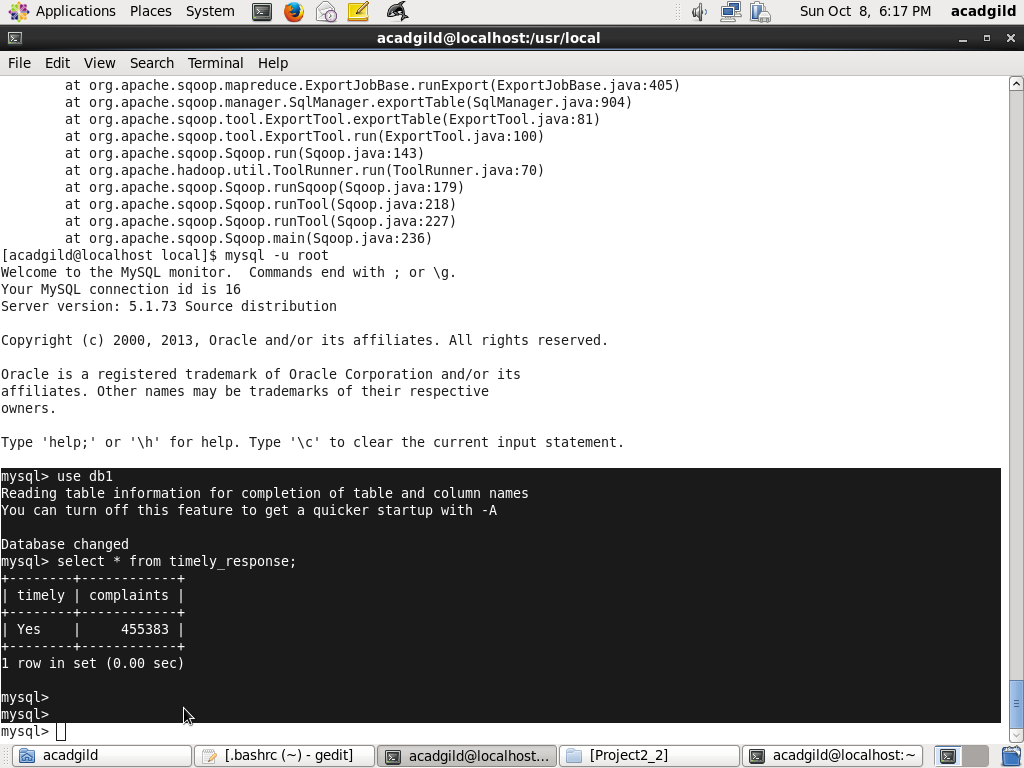
Store Pig Latin output to HDFS

Store F into ‘/output/first’ using PigStorage(‘,’);

Sqoop export--------------------



Store in database------------------------



**Task 3**

**Write a pig script to find no of complaints where consumer forum forwarded the complaints same day they received to respective company**

A = load '/user/acadgild/Project2/Consumer\_Complaints.csv' using CSVLoader(',') as(f1:chararray, f2:chararray, f3:chararray , f4:chararray, f5:chararray, f6:chararray, f7:chararray, f8:chararray, f9:chararray, f10:chararray,f11:chararray,f12:chararray,f13:chararray,f14:chararray,f15:chararray,f16:chararray);

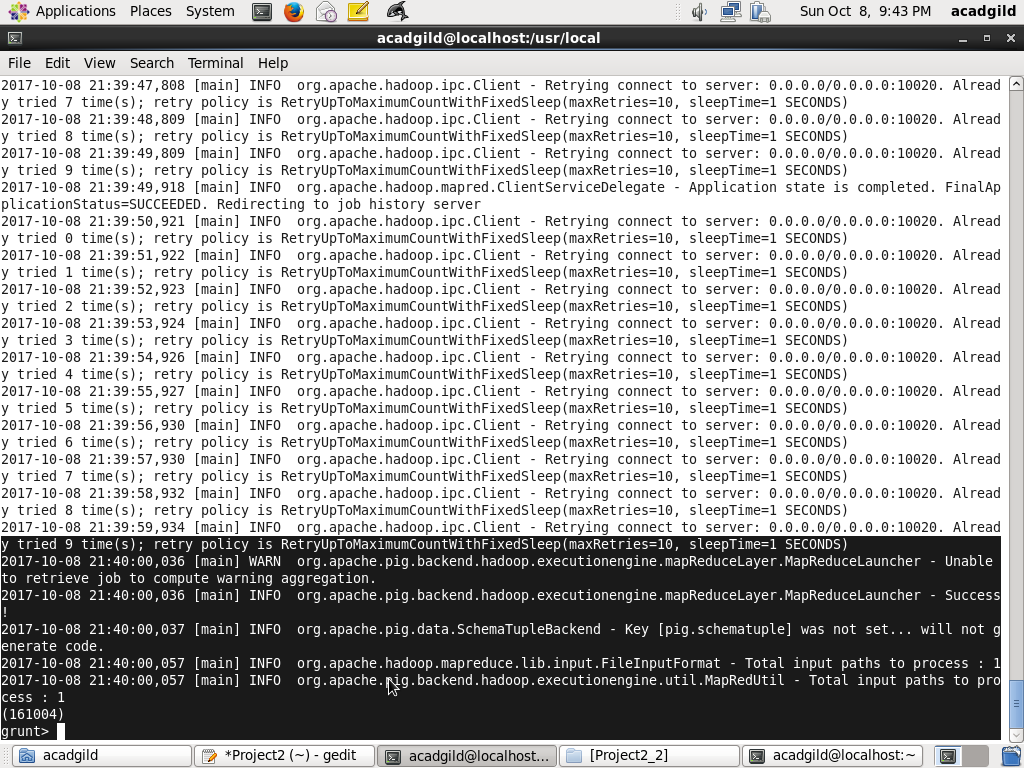
B = foreach A generate f1 as compdate,f12 as forwdate,f16 as id;

C =filter B by compdate==forwdate ;

D = group C all;

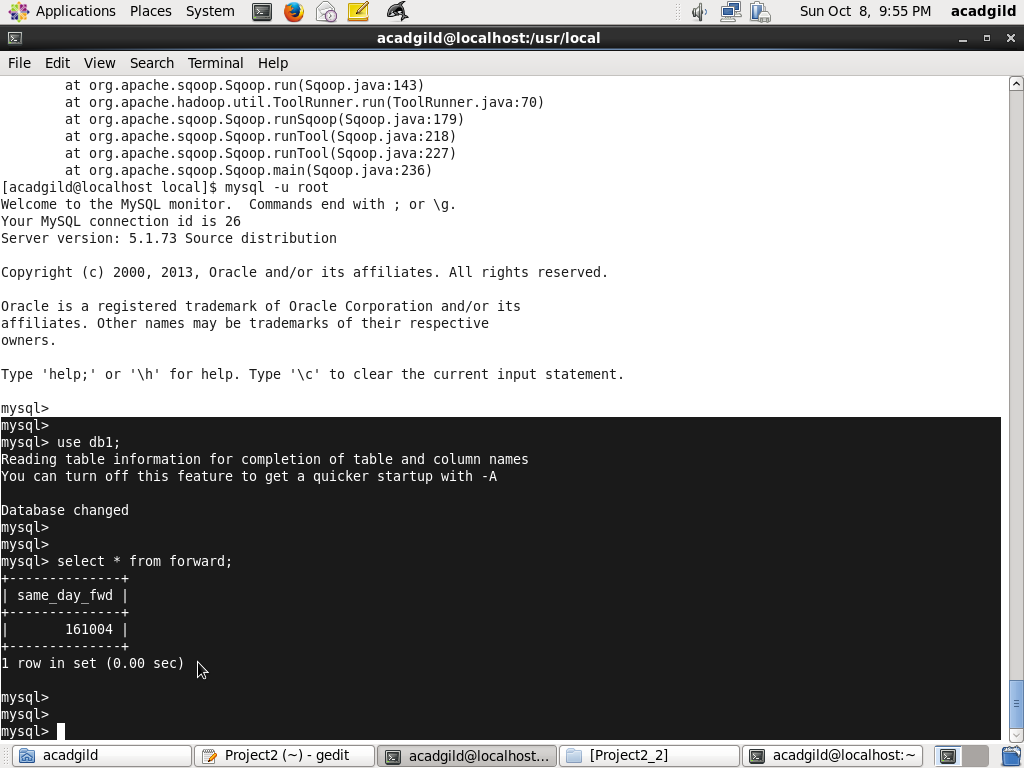
E= Foreach D generate COUNT(D.id)as same\_day\_forward;

Dump E;



store E into '/output/second' using PigStorage(',')

Sqoop export --connect jdbc:mysql://localhost/db1 –username root –table forward -m 1 – export-dir ‘/output/second/’



**Task 4**

**Write a pig script to find list of companies toping in complaint chart (companies with maximum number of complaints)**

A = load 'user/acadgild/Project2/Consumer\_Complaints.csv' using CSVLoader(',');

B = foreach A generate $7 as company,$15 as id;

C = group B by company;

D = foreach C generate REPLACE(group,',',' '),COUNT(B.id) as numofcomplaints;

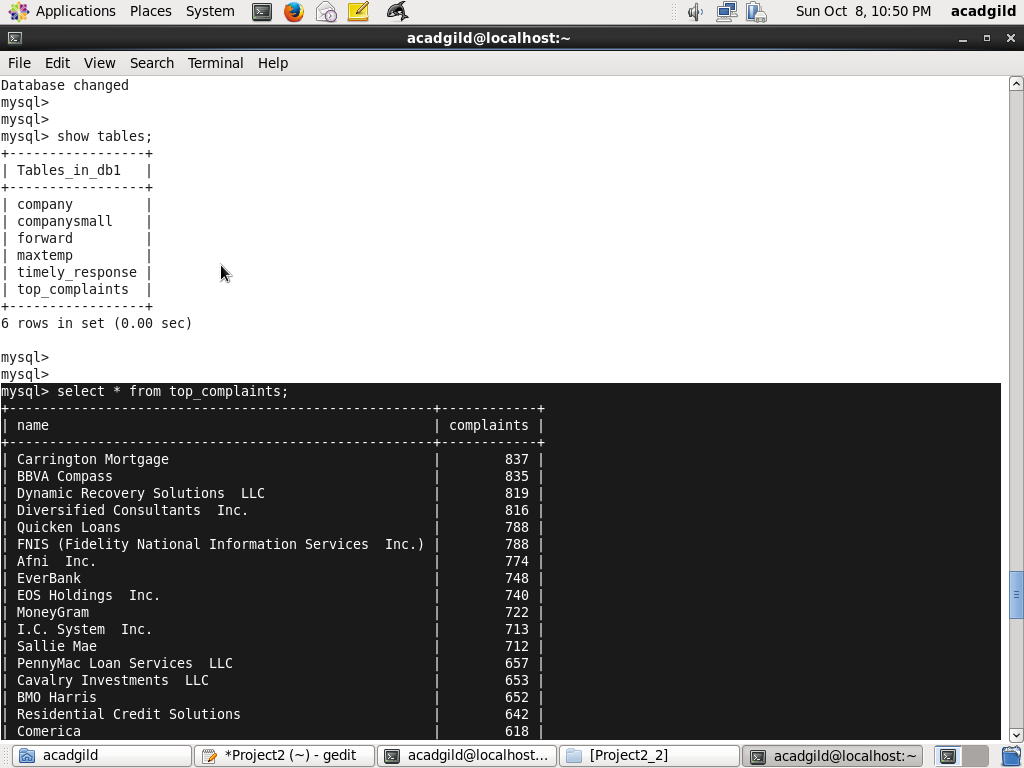
E = Order D by numofcomplaints desc;

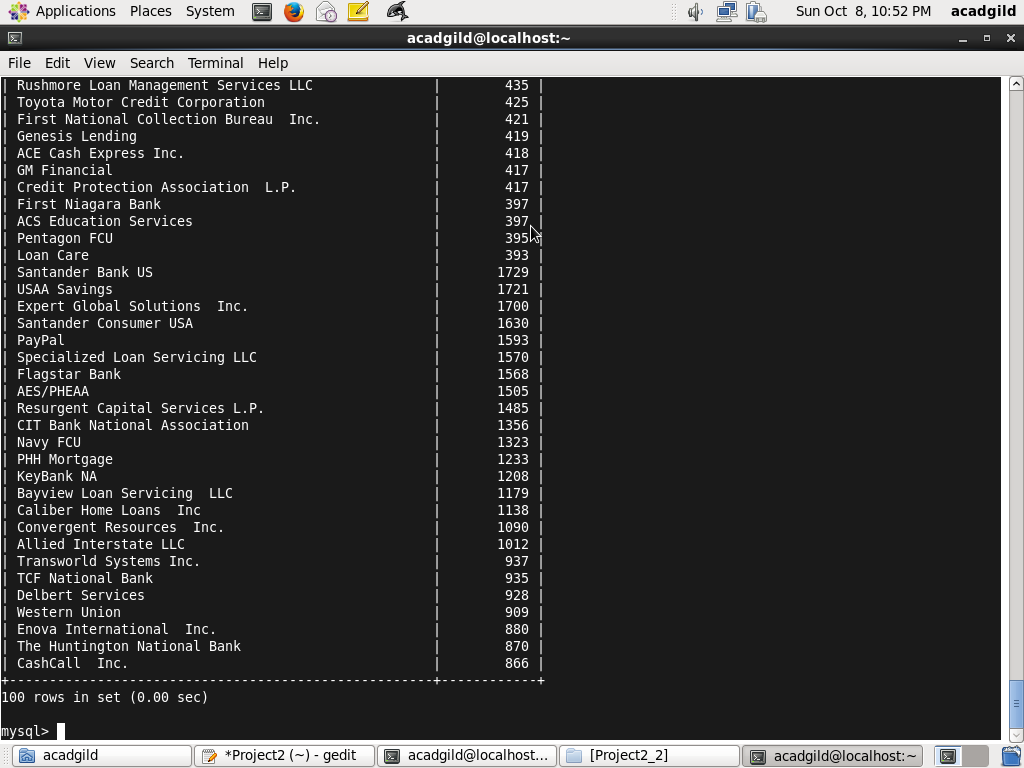
F = limit E 100;

Dump F;

Store F into ‘/output/Third’ as PigStorage(‘,’)

Sqoop export --connect jdbc:mysql://localhost/db1 –username root –table top\_complaints -m 1 –export-dir ‘/output/Third/’





**Task 5**

**Write a pig script to find no of complaints filed with product type has “ Debt collection” for the year 2015**

A = load '/user/acadgild/Project2/Consumer\_Complaints.csv' using CSVLoader(',') as (f1:chararray,f2:chararray);

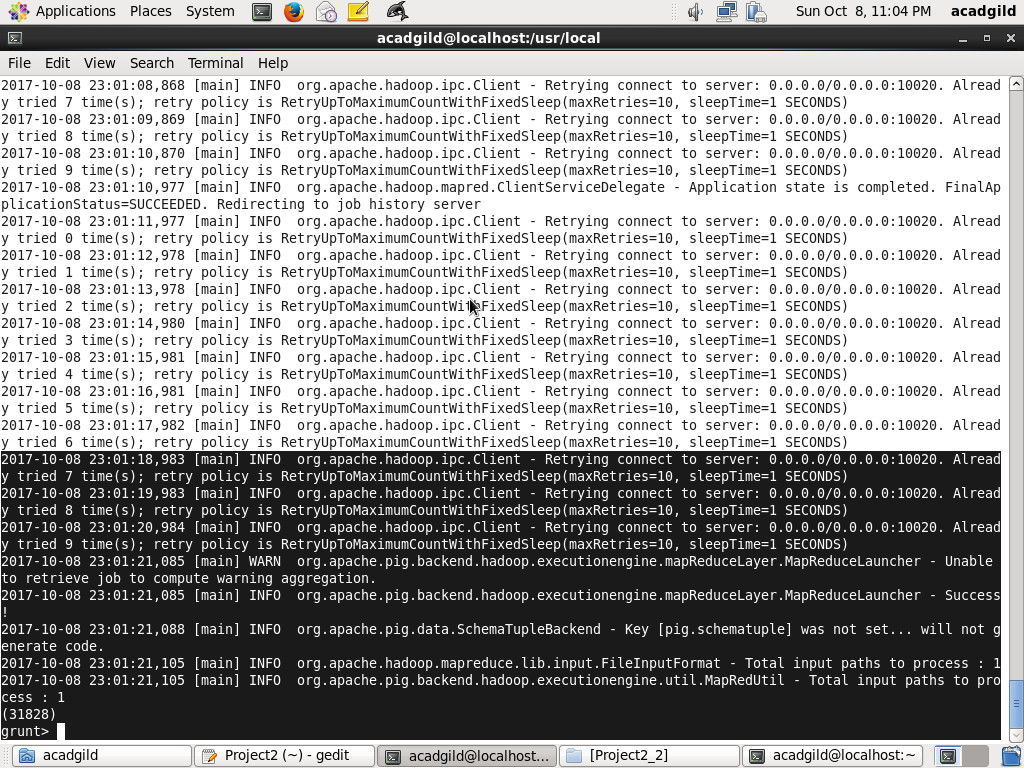
B = filter A by f2=='Debt collection' and GetYear(ToDate(f1,'MM/dd/yyyy'))==2015;

C = foreach B generate f1 as complaintDate,f2 as product;

D = group C all;

E = Foreach D generate COUNT(C.product);

Dump E;



Store E into '/output/Four' as PigStorage(',');

Sqoop export --connect jdbc:mysql://localhost/db1 --username 'root' --table 'complaints' --export-dir '/output/Four/'

