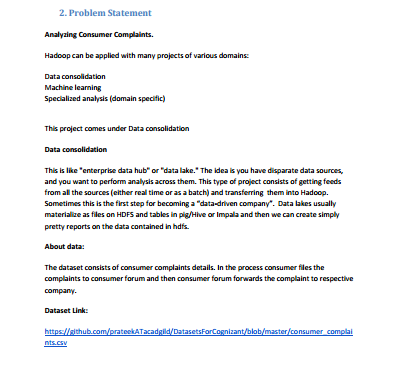
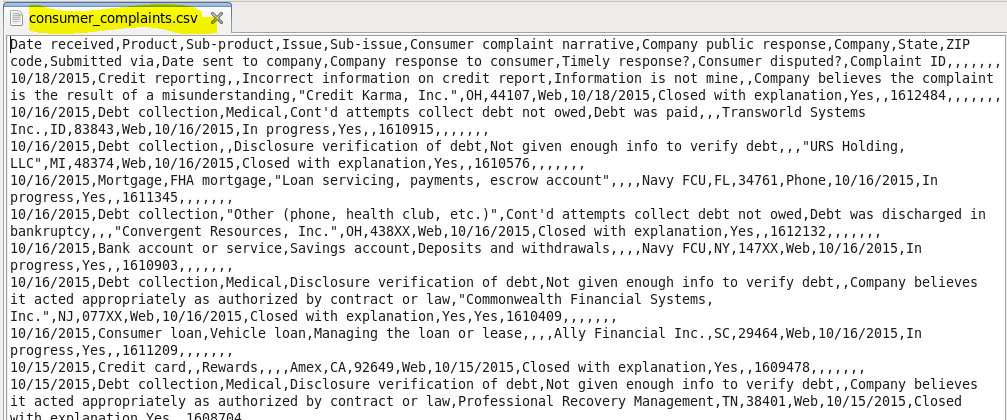
**#PROJECT-2**

**Assignment\_13.2**

****

**Data set- ’consumer\_complaints.csv’**

****

**PROBLEM STATEMENT-1**

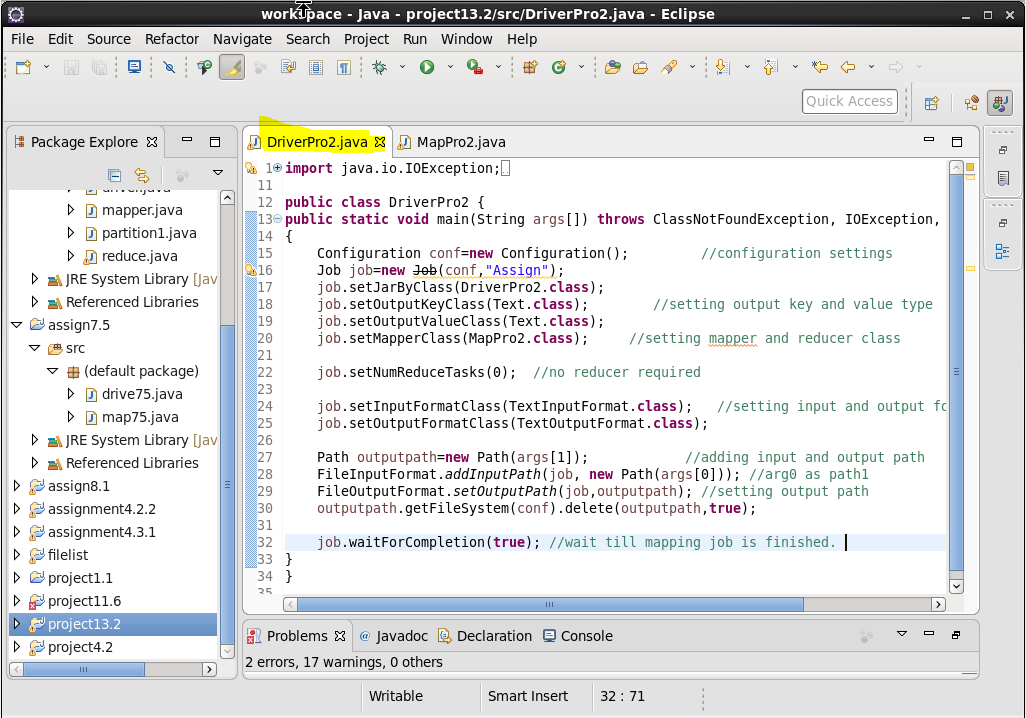
**Write a mapreduce program to remove commas present inside the double quotes.**

**PROGRAM:**

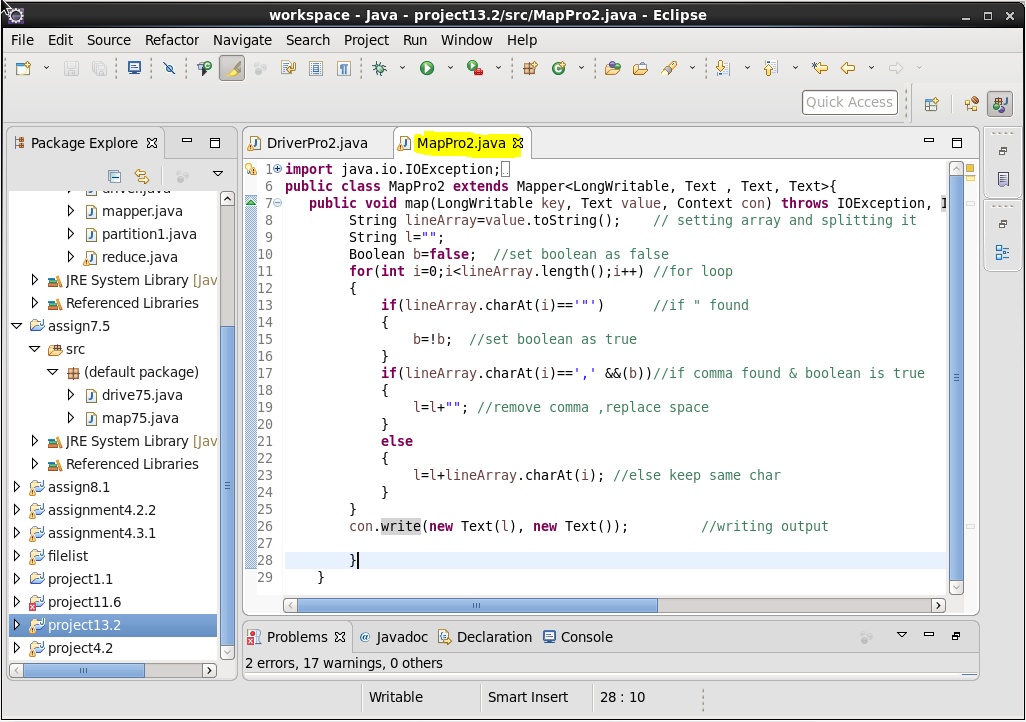
We need to remove the commas that are present inside the double quotes.

To do that we will first check occurrence of double quote and will later replace by space if comma is found inside the data set **‘consumer\_complaints.csv’**.

**DRIVER CLASS:**

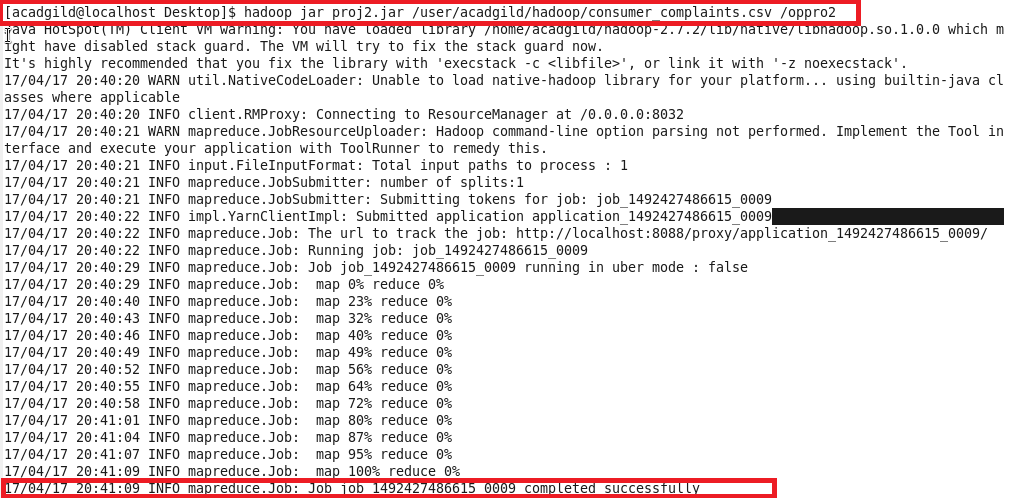
****

**MAPPER CLASS:**

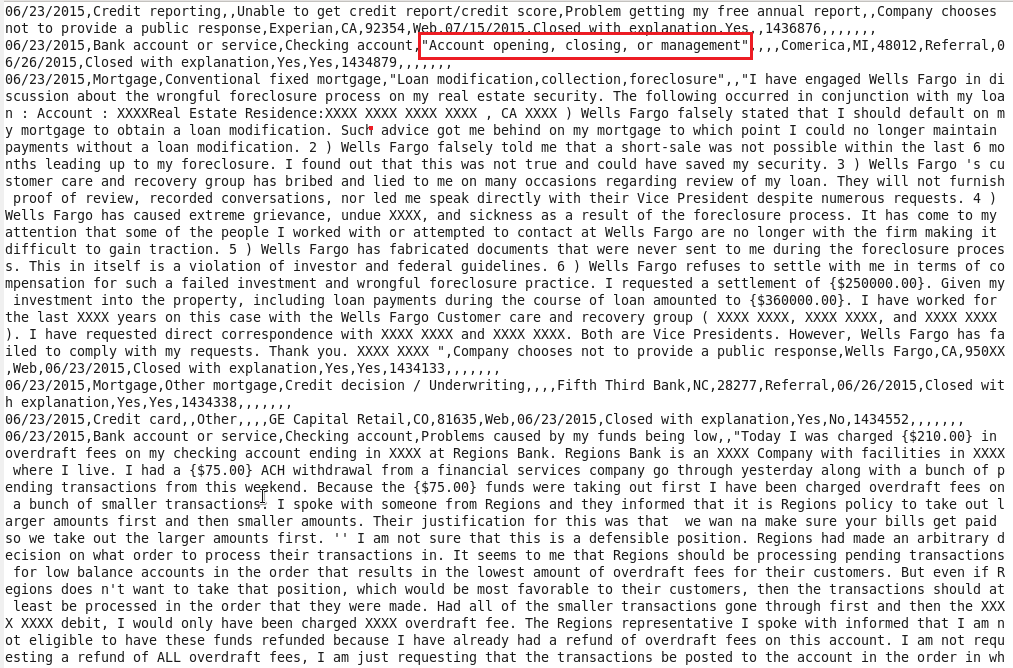
****

**RESULT:**

**Loading Jar file-**

****

**Input file:**

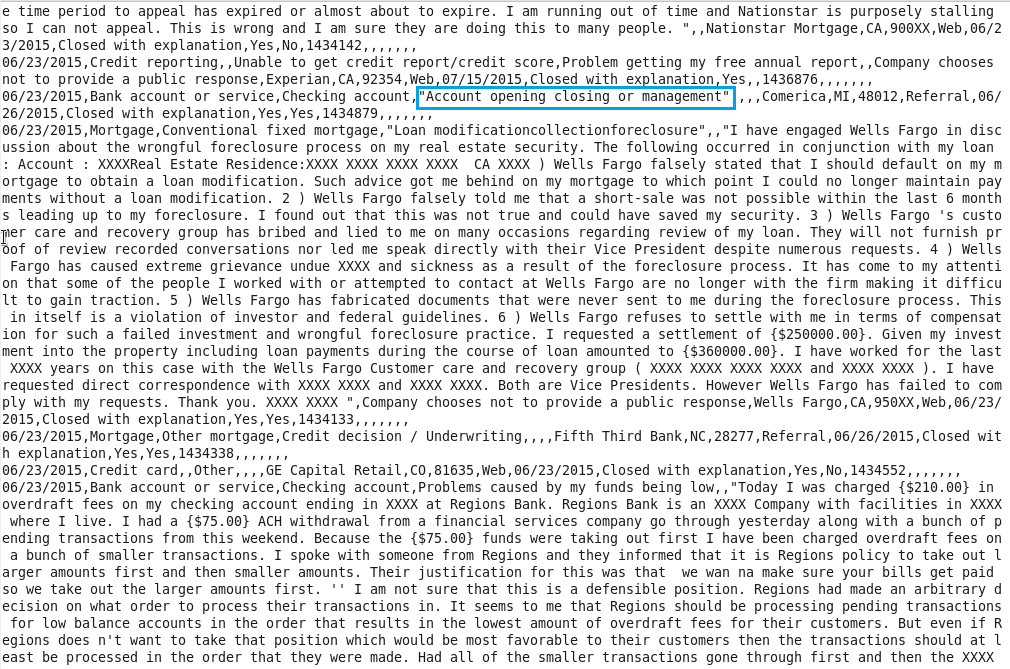
****

Let’s consider part of input file consumer\_complaints.csv to verify the result.

**“Account opening, closing, or management”**

When we run the jar file the commas present inside are removed successfully. The result is attached below

**Output:**

****

Now we have to use this output file as input file for next pig based questions.

This output file is saved as **‘consume.txt’**

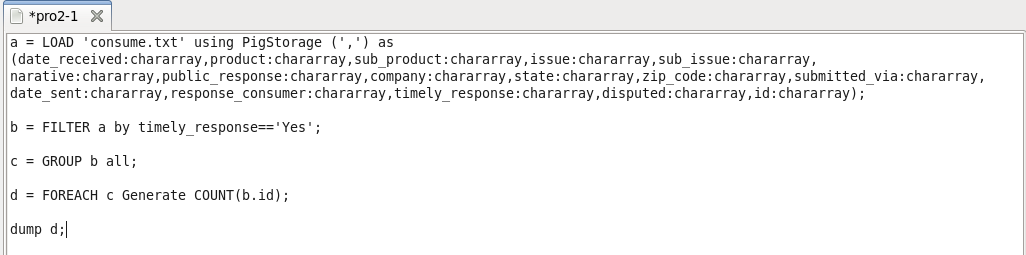
**Problem statement-2.1**

Write a pig script to find no of complaints which got timely response

**PigScript:**

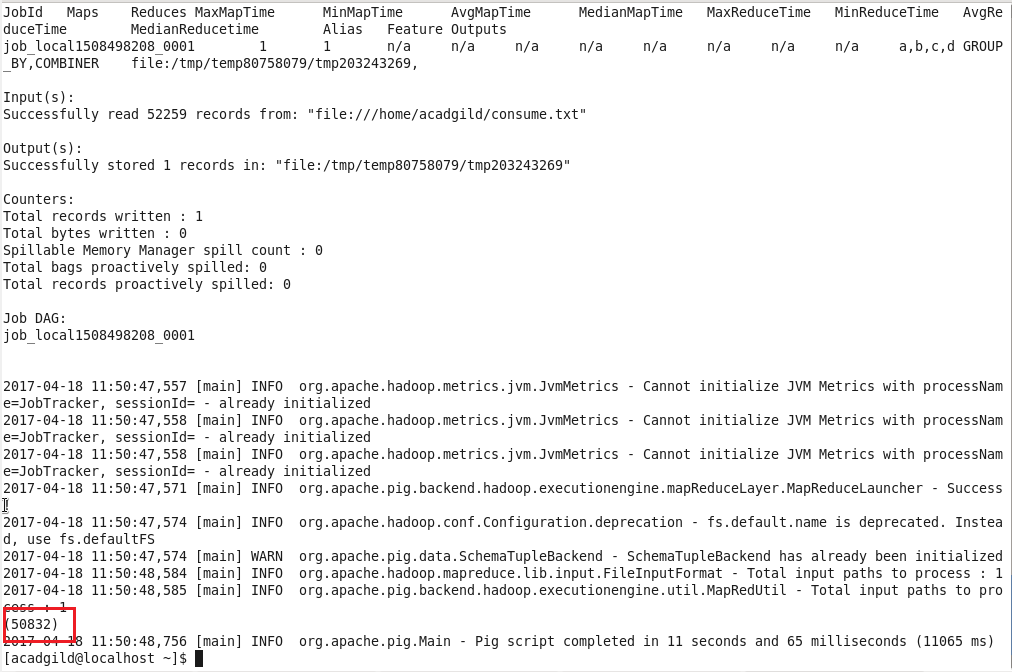
Creating gedit file and writing pig commands inside

First we will load the file by defining column name and data type, will then filter according to requirement like timely\_response as yes and will then group and count the ids having timely response as yes.

****

**Result:**

Running the gedit file and display result on console

****

In all total **50832** ids has timely response as yes.

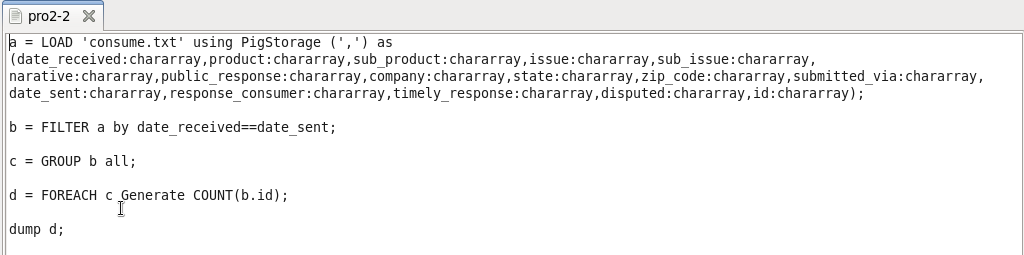
**Problem statement-2.2**

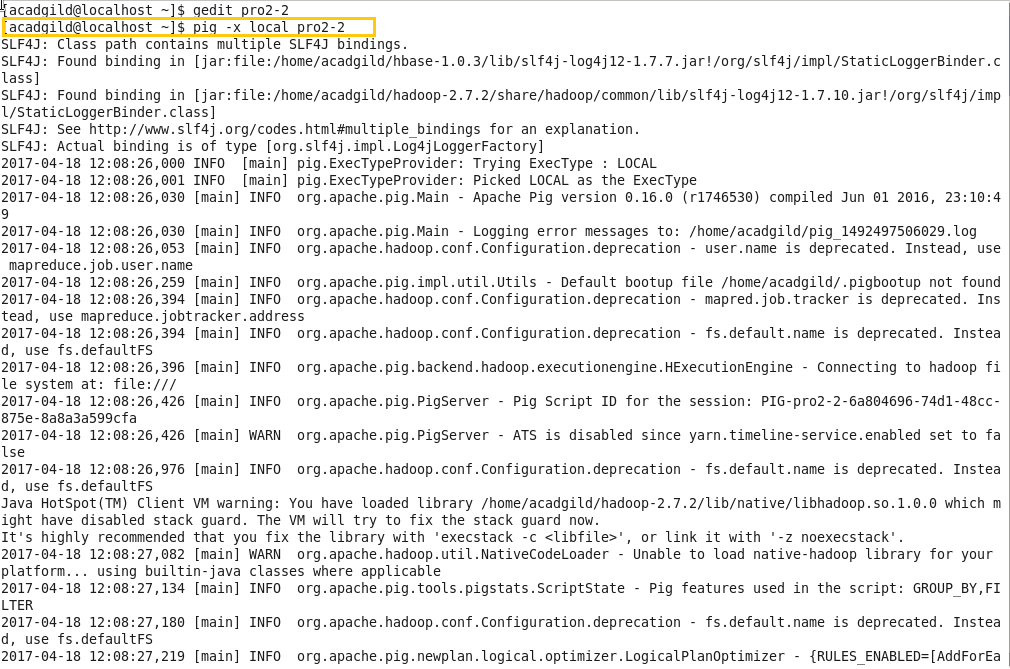
Write a pig script to find no of complaints where consumer forum forwarded the complaint same day they received to respective company

**PigScript:**

Creating gedit file and writing pig commands inside

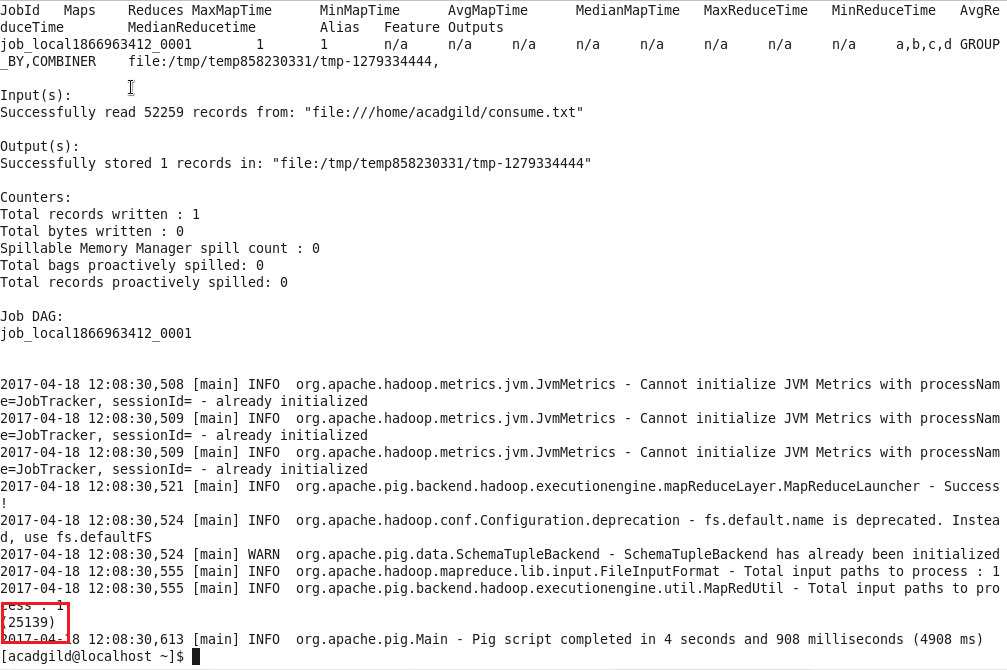
First we will load the file by defining column name and data type, will then filter according to requirement like date of sent is same as date received and will then group and count the ids having same dates.

****

****

**Result:**

Running the gedit file and display result on console

****

No of complains which were sent and received on same day are 25139.

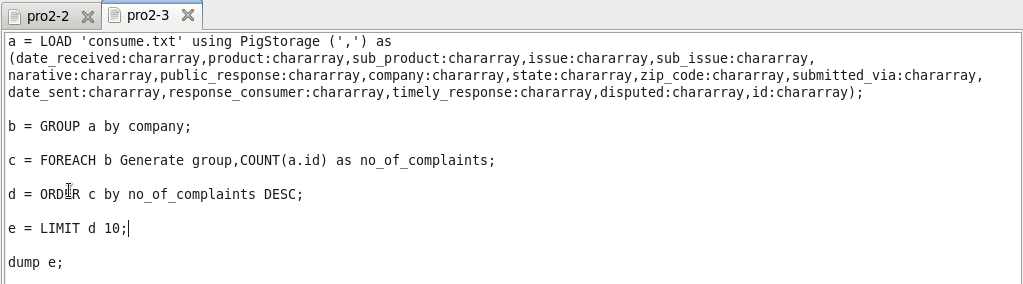
**Problem statement-2.3**

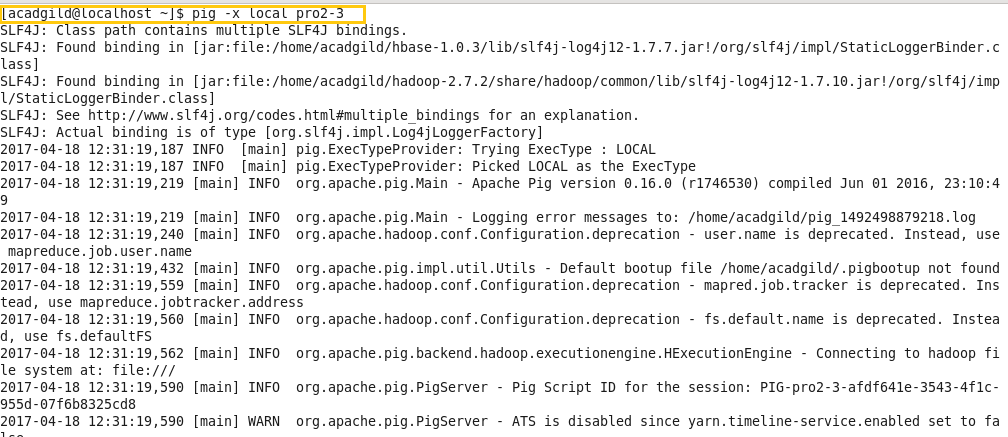
Write a pig script to find list of companies topping in complaint chart (companies with maximum number of complaints)

**PigScript:**

Creating gedit file and writing pig commands inside

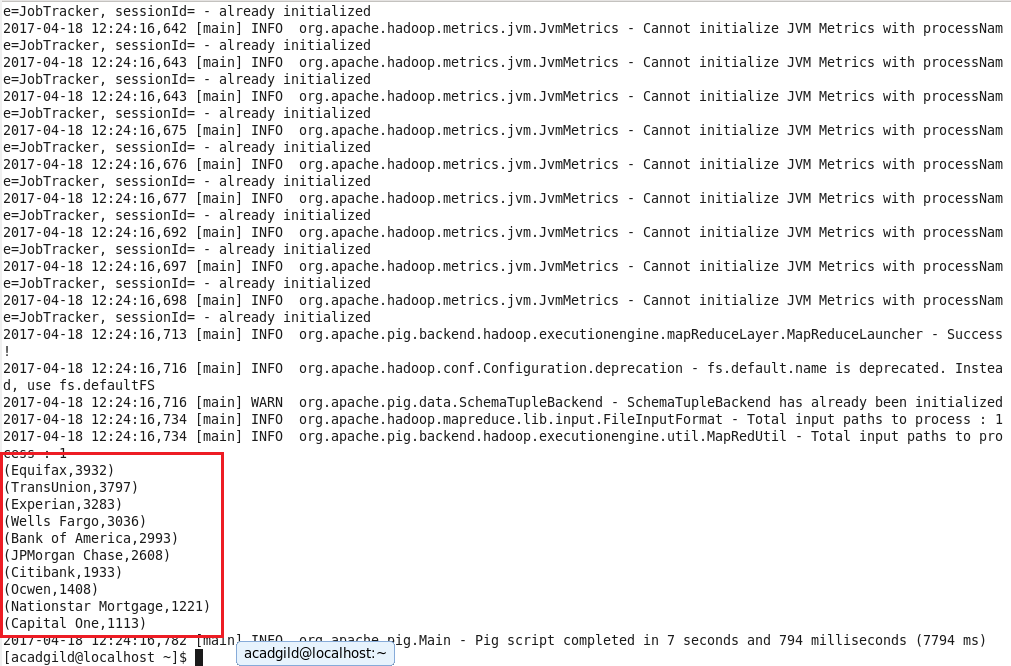
First we will load the file by defining column name and data type, will then filter according to requirement like count complain counts and will then group and count the ids having same and limit result by 10.

****

****

**Result:**

Running the gedit file and display result on console

****

Top 10 results of companies having max. Employees.

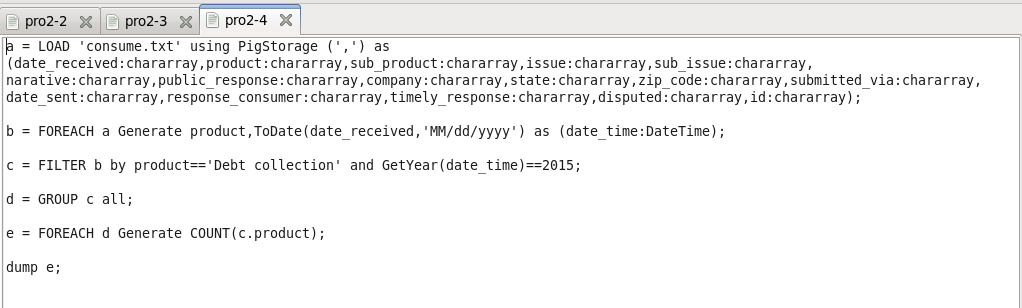
**Problem statement-2.4**

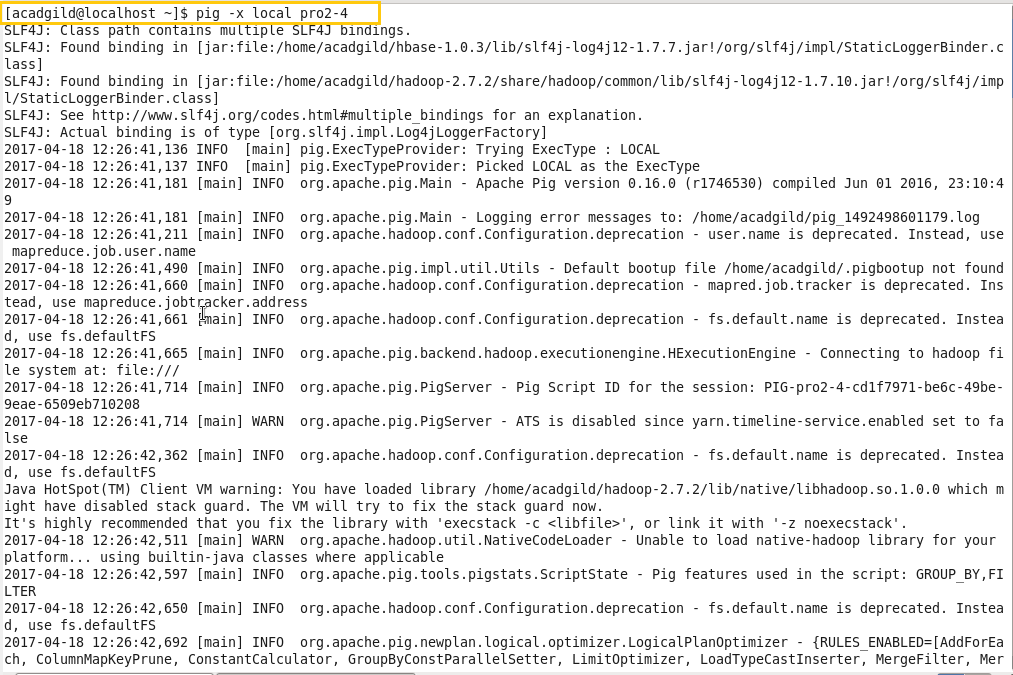
Write a pig script to find no of complaints filed with product type has "Debt collection" for the year 2015

**PigScript:**

Creating gedit file and writing pig commands inside

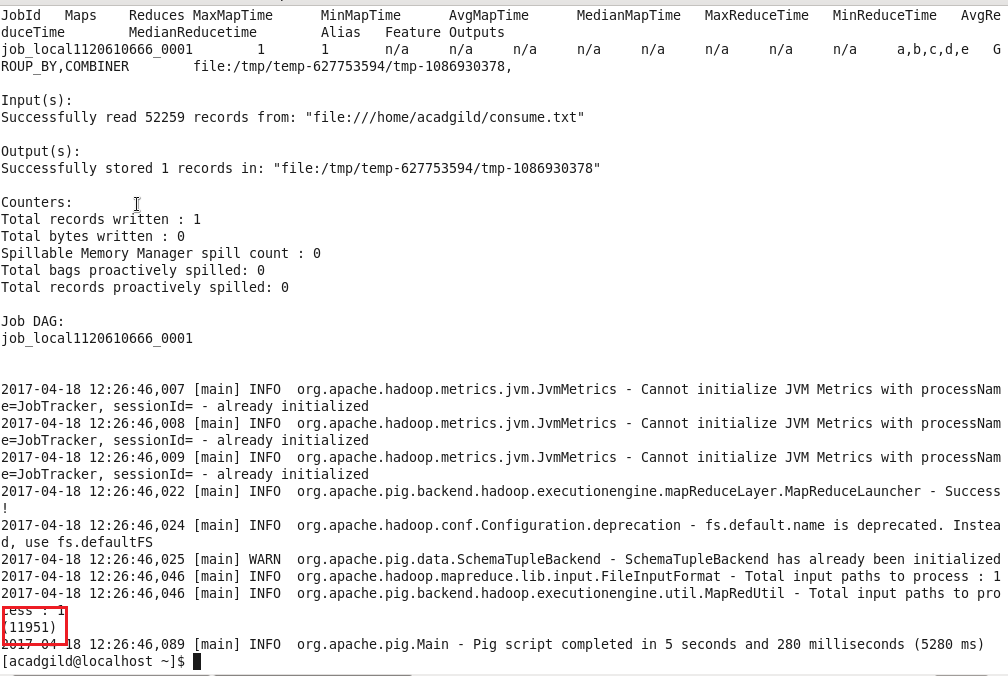
First we will load the file by defining column name and data type, will then filter according to requirement like complain is debt collection and year as 2015 ,will then group and count the ids having same and limit result by 10

****

****

**Result:**

Running the gedit file and display result on console

****

We can see the count of complains of 2015 in debt collection as 11951.