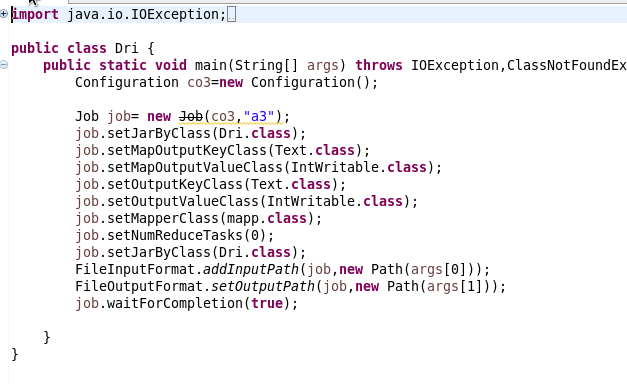
**Project-2**

**Project 2 - Analyzing Consumer Complaints**

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

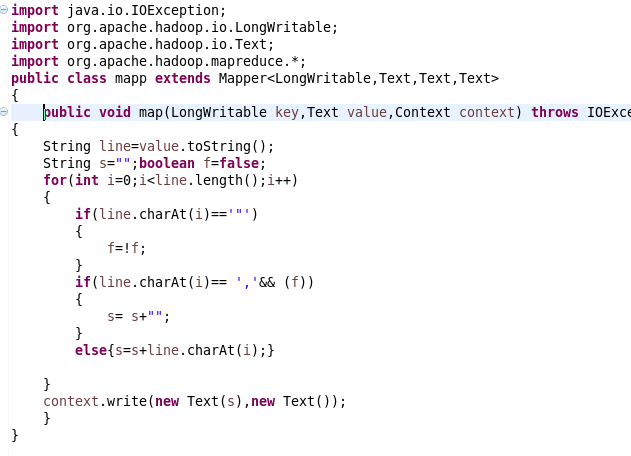
The question is that the input data that we are taking in tokenize by “,” but when there is a comma that does not describe a separate column but the content value.So to distinguish this ,we are first checking the double quotes and the content inside the double quotes is comma ,if it is so replace it by space .So that it does not get any confusion on dividing columns

Driver class: At first we are giving the driver class and mapper class and the input file path ,output file path .



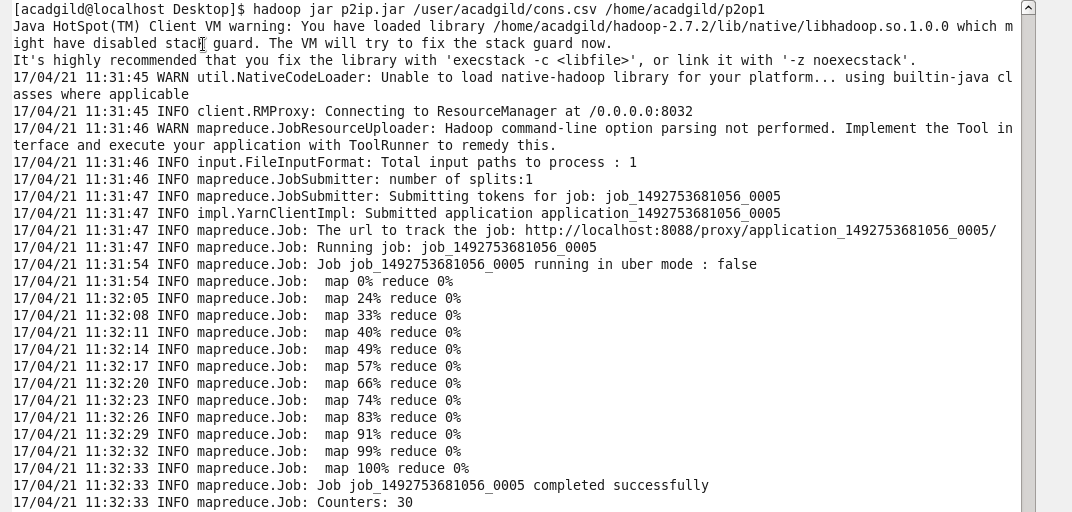
Mapper Class: In mapper class ,first we are checking for double quotes. If it is a double quotes then we are setting the Boolean value into true ,so following this the contents are compared with comma .If it is comma ,then it is replaced by space ,if not then it moves to the next character.

To distinguish the opening quote and losing one,we are setting Boolean value as true for opening and it is setting false for closing.So that if both comma and boolean value satisfies,then it replace the content.

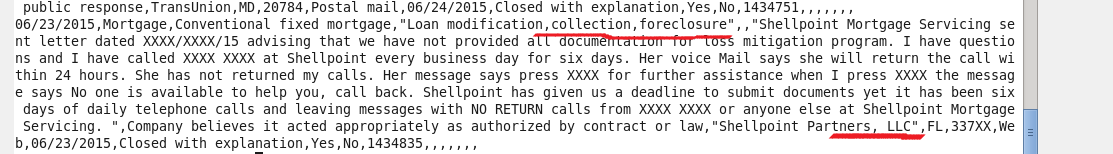


Now the output of this program is used as input for the following program.

Run:

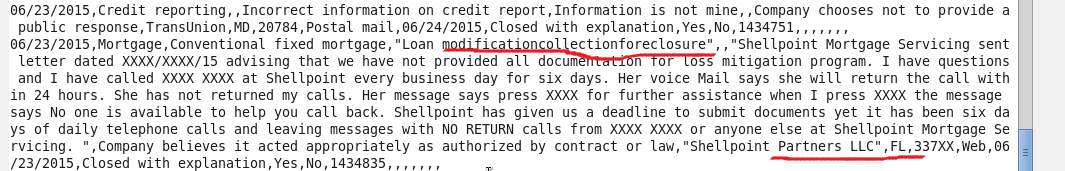


Output terminal:

Without removing comma: 

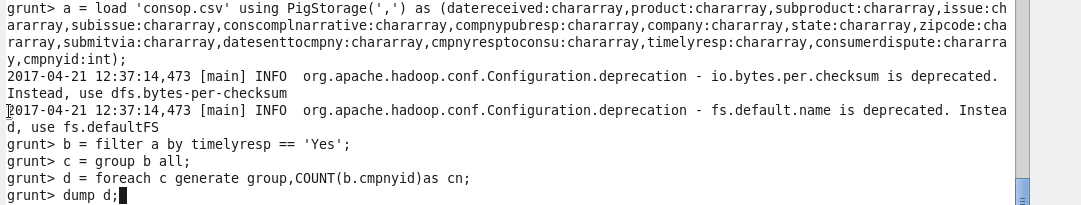
After removing comma:

asdsfddf



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

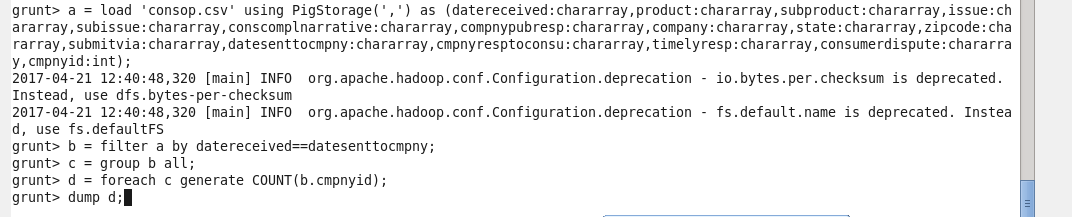
**Pigscript:** Here we need no of complaints that got timely response,so we are filtering based on timely response = yes ,then we are counting the no of complaints.

****

**Output terminal:** **C:\Users\612912\Pictures\pro 2.1 op.PNG**

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

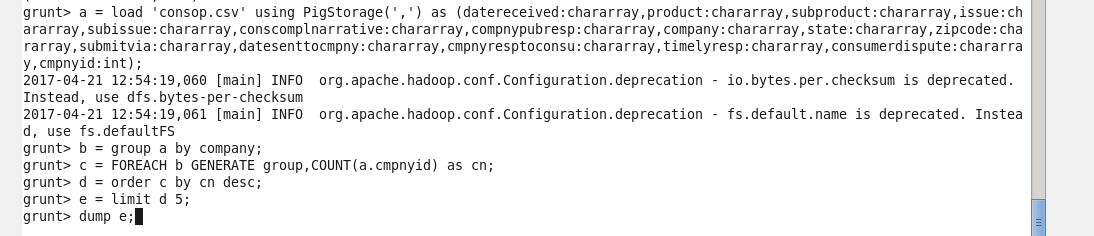
**Pigscript:** Herewe are checking whether the complaint submitted date and then the complaint received date is same then we are counting the no of complaints.

****

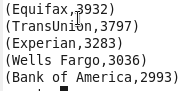
**Output terminal** **C:\Users\612912\Pictures\proj 2.2 op.PNG**

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

**Pigscript:** In thiswe are grouping by company and then the complaints are counted. Then ordered in descending order .Here I am displaying only top 5 companies, so I am setting the limit as 5

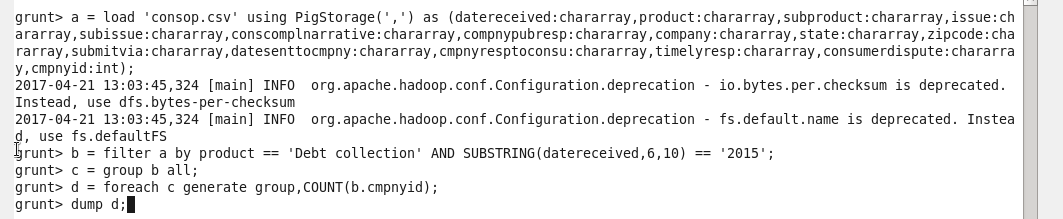


**Output terminal**



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

**Pigscript:** Here we are filtering the product type to be Debt collection and from that the complaint date received to be in the year of 2015 by getting the substring of date that corresponds to year only .Then we are counting the complaint id to get the no of complaints.

****

**Output terminal** **C:\Users\612912\Pictures\proj 2.4 op.PNG**