1. **Introduction**

This assignment will help you to consolidate the concepts learnt in the session.

1. **Problem Statement**

Analyzing Consumer Complaints.

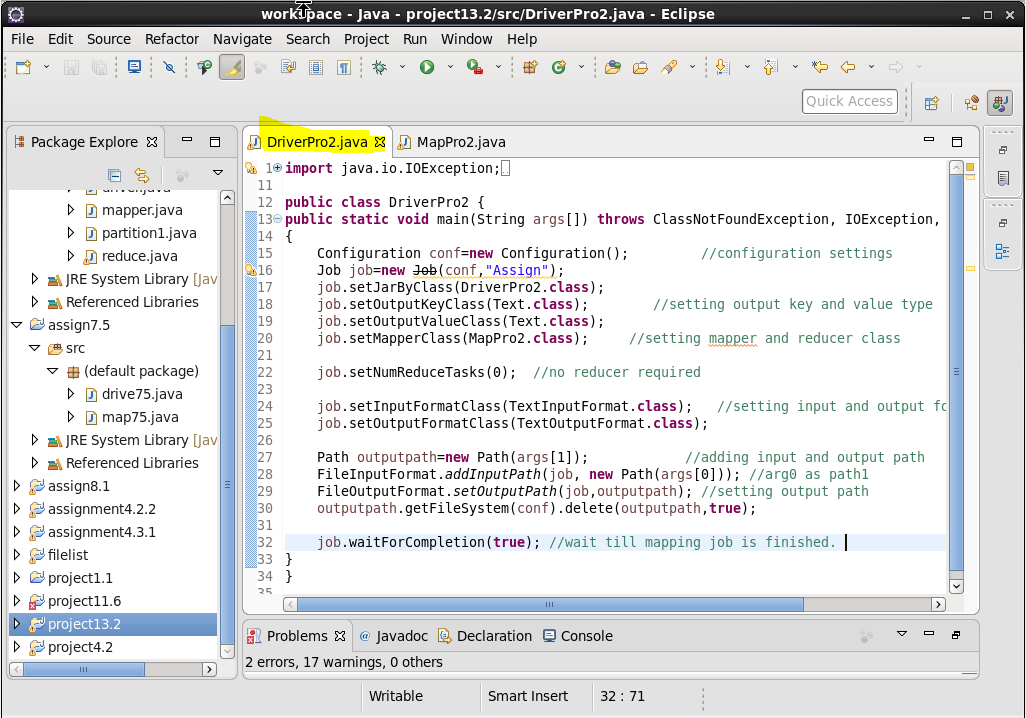
The dataset consists of consumer complaints details. In the process consumer files the complaints to consumer forum and then consumer forum forwards the complaint to respective company.

**Dataset Link:** [https://github.com/prateekATacadgild/DatasetsForCognizant/blob/master/consumer\_complai nts.csv](https://github.com/prateekATacadgild/DatasetsForCognizant/blob/master/consumer_complai%20nts.csv)

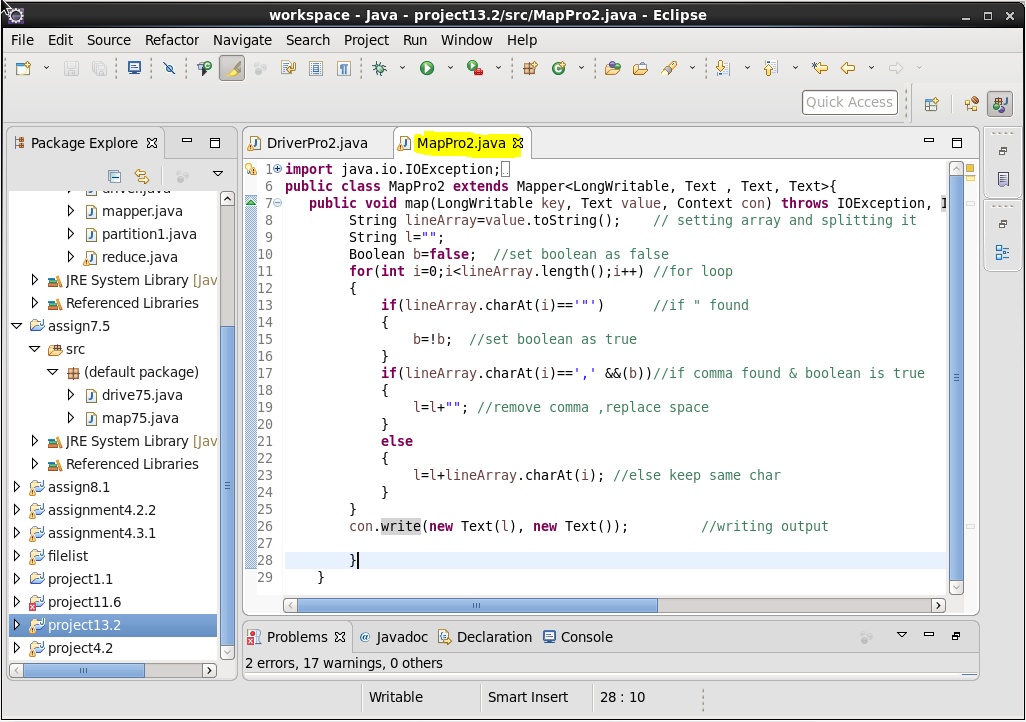
1. **Write a mapreduce program to remove commas present inside the double quotes.**

Note: Work on this problem statements after doing the data cleaning as mentioned above.

Driver-

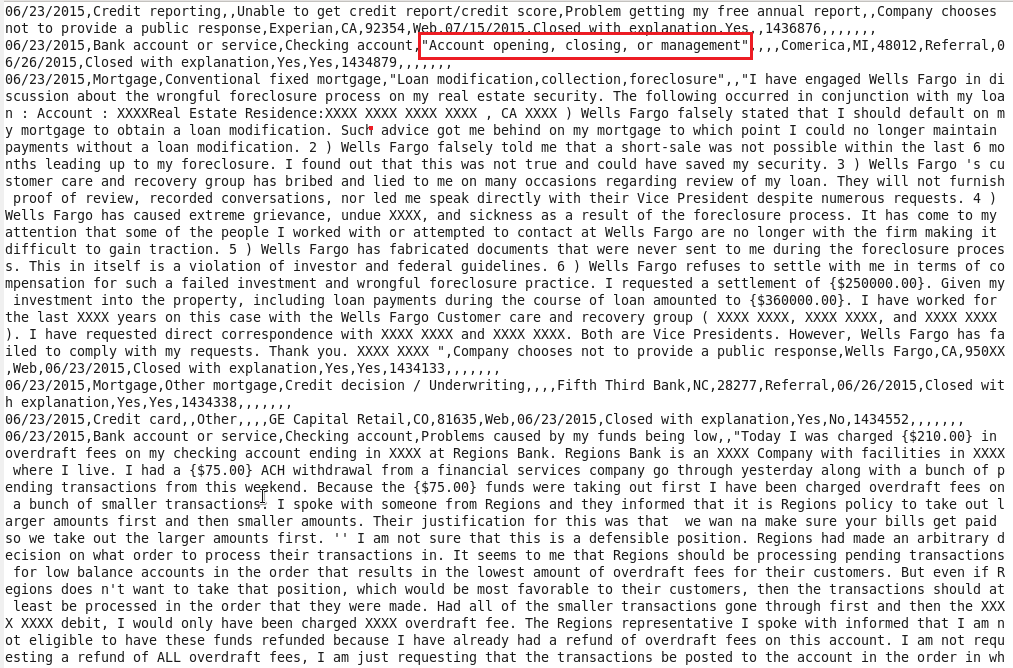


Mapper-

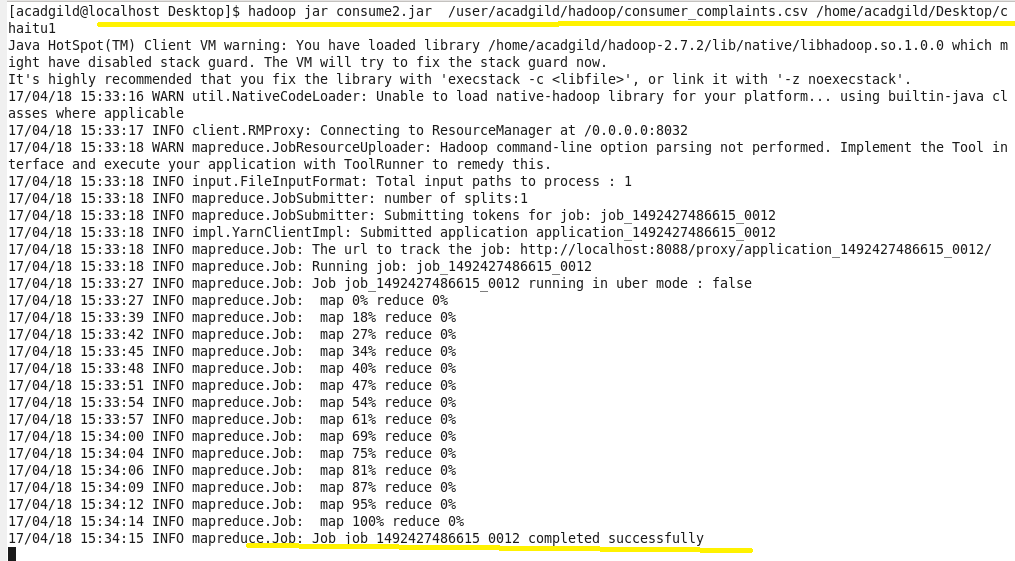


Result:

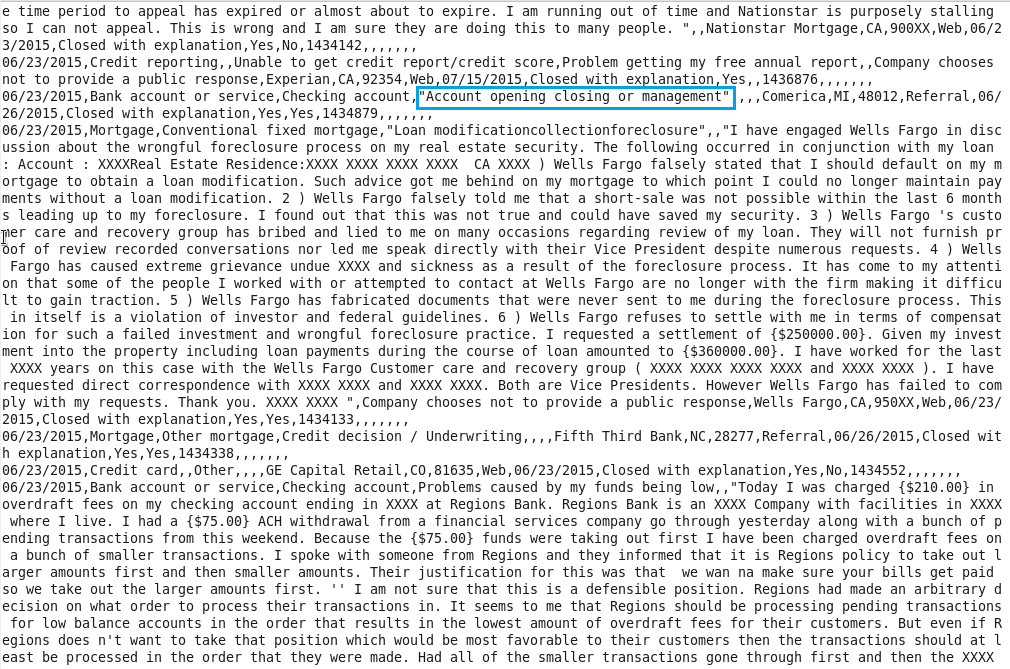
Input file



Load jar file



Result-

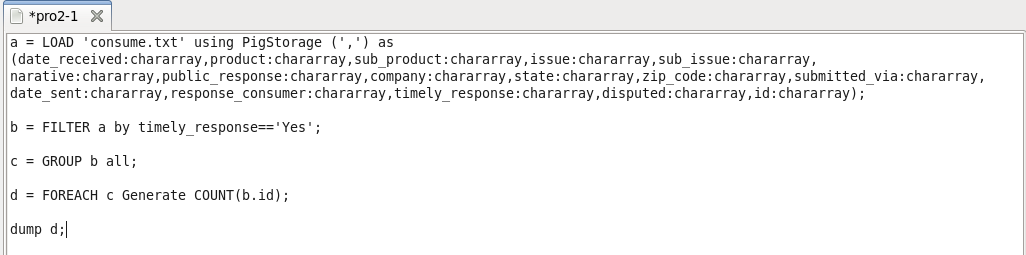


The commas are removed from double quotes.

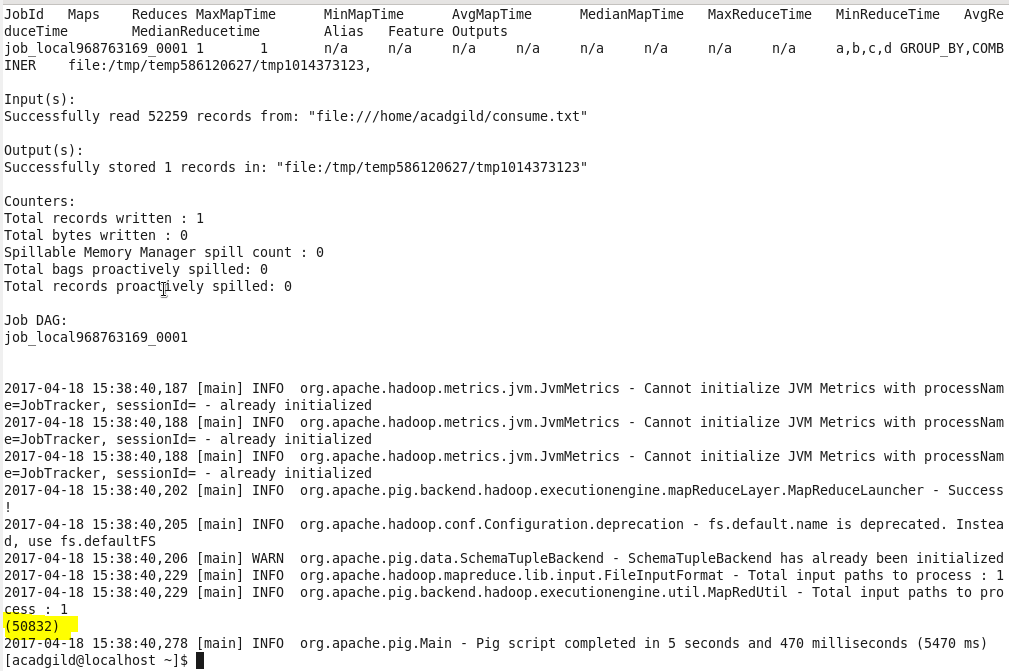
Pig Scripting-

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

Pig command-

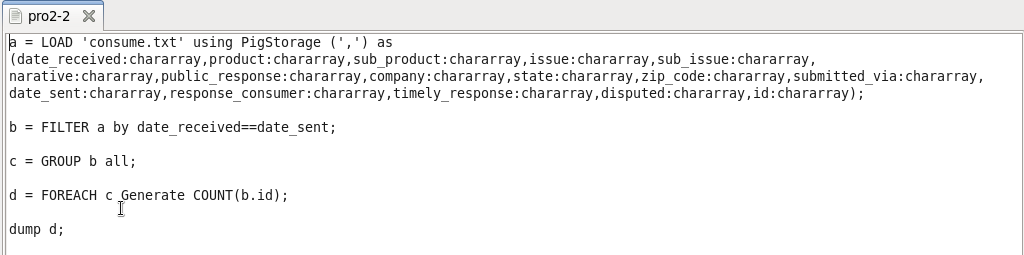


Result**-**

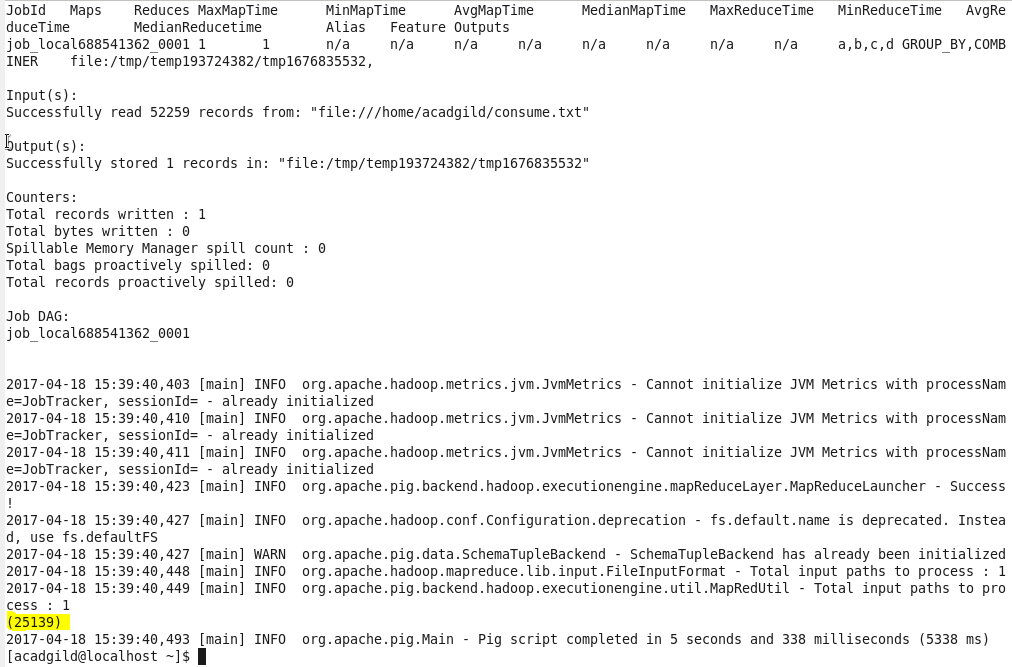
****

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

Pig command-

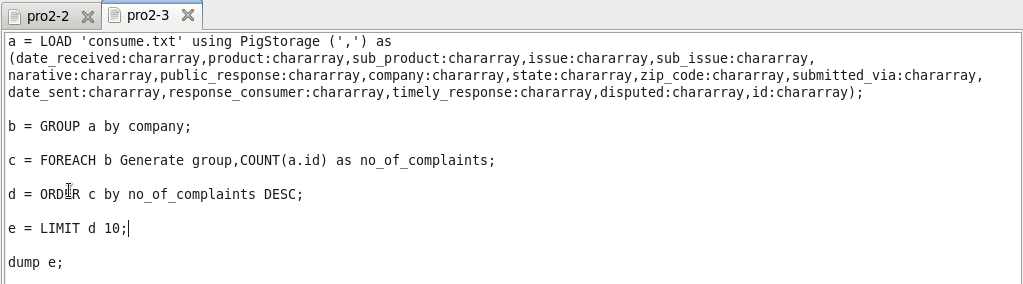


Result-

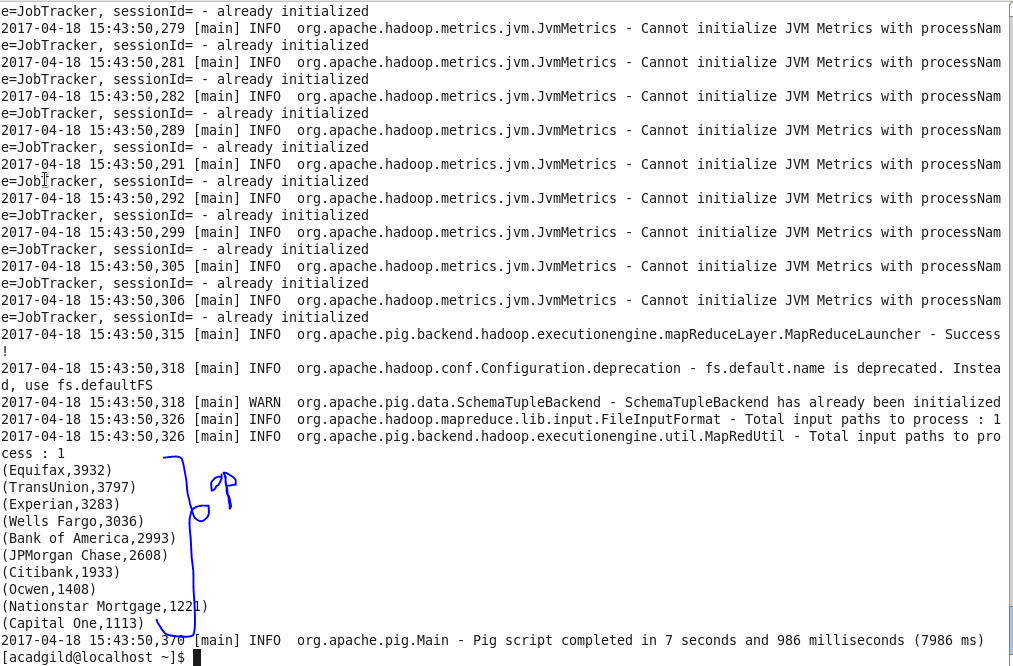


1. **Write a pig script to find list of companies topping in complaint chart (companies with maximum number of complaints)**

Pig command-

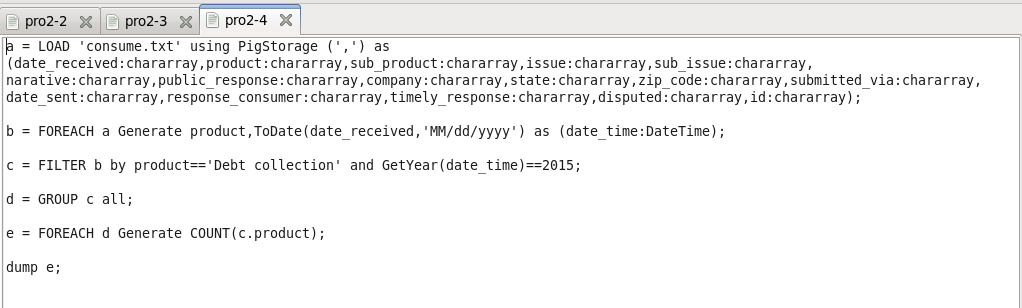


Result

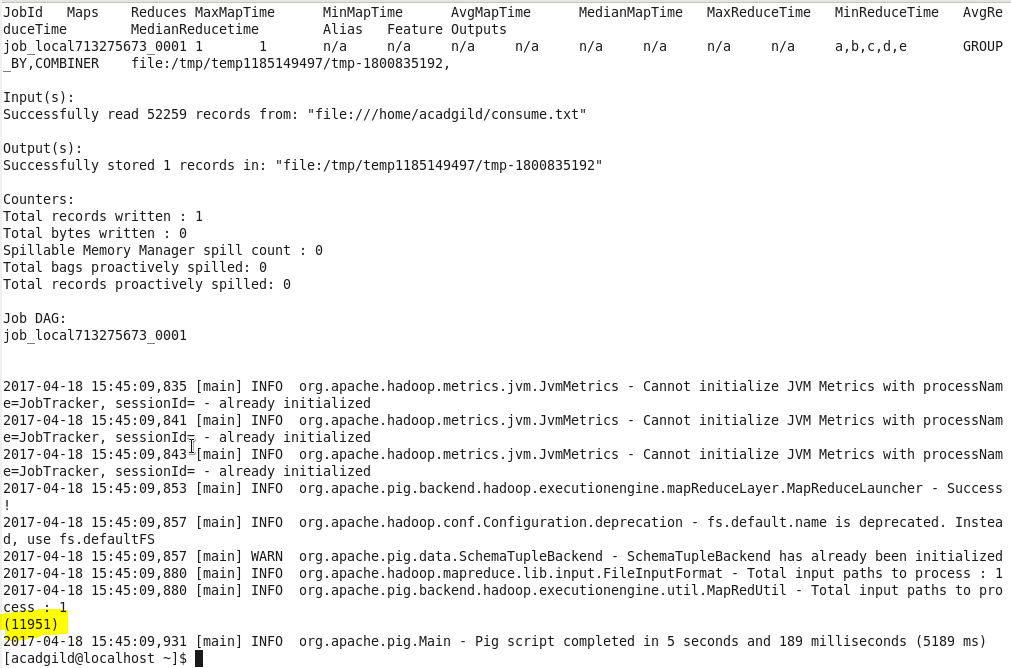
****

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

Pig command-



Result-

****