**PROBLEM STATEMENT:**

Analyzing Consumer Complaints.

Hadoop can be applied with many projects of various domains:

Data consolidation Machine learning Specialized analysis (domain specific) This project comes under Data consolidation Data consolidation This is like "enterprise data hub" or "data lake." The idea is you have disparate data sources, and you want to perform analysis across them. This type of project consists of getting feeds from all the sources (either real time or as a batch) and transferring them into Hadoop. Sometimes this is the first step for becoming a “data-driven company”. Data lakes usually materialize as files on HDFS and tables in pig/Hive or Impala and then we can create simply pretty reports on the data contained in hdfs.

About data: 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.

**I.**

**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.**

**II.**

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

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

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

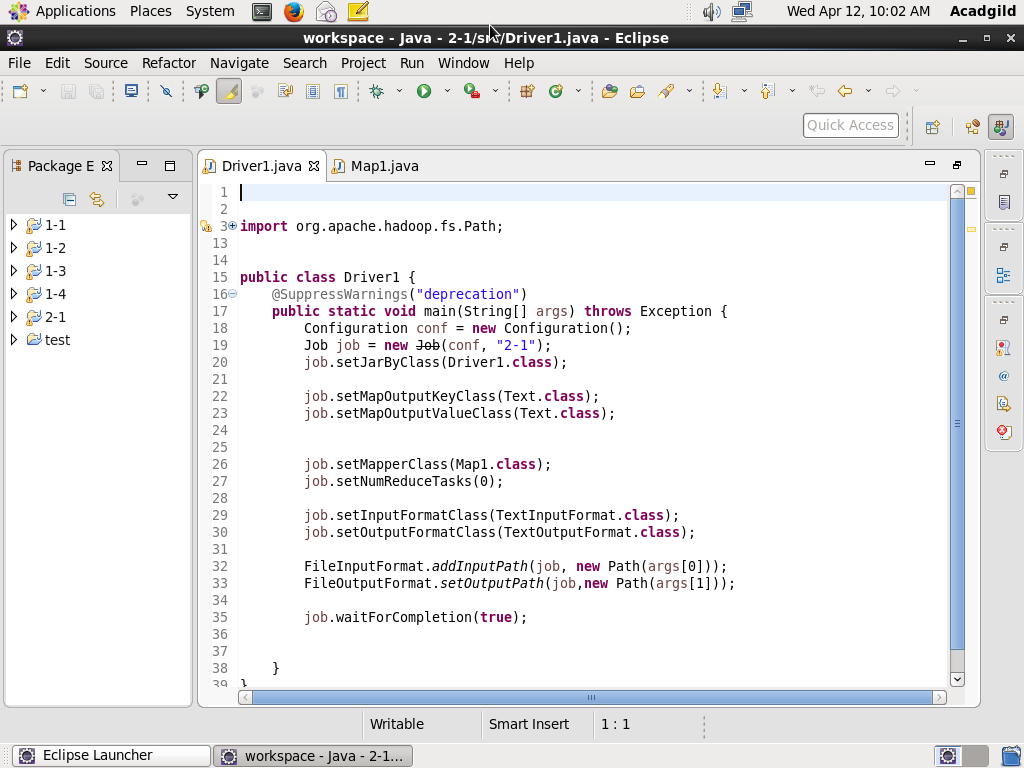
**4. Write a pig script to find no of complaints filed with product type has "Debt collection" for the year 2015**

**SECTION-1:**

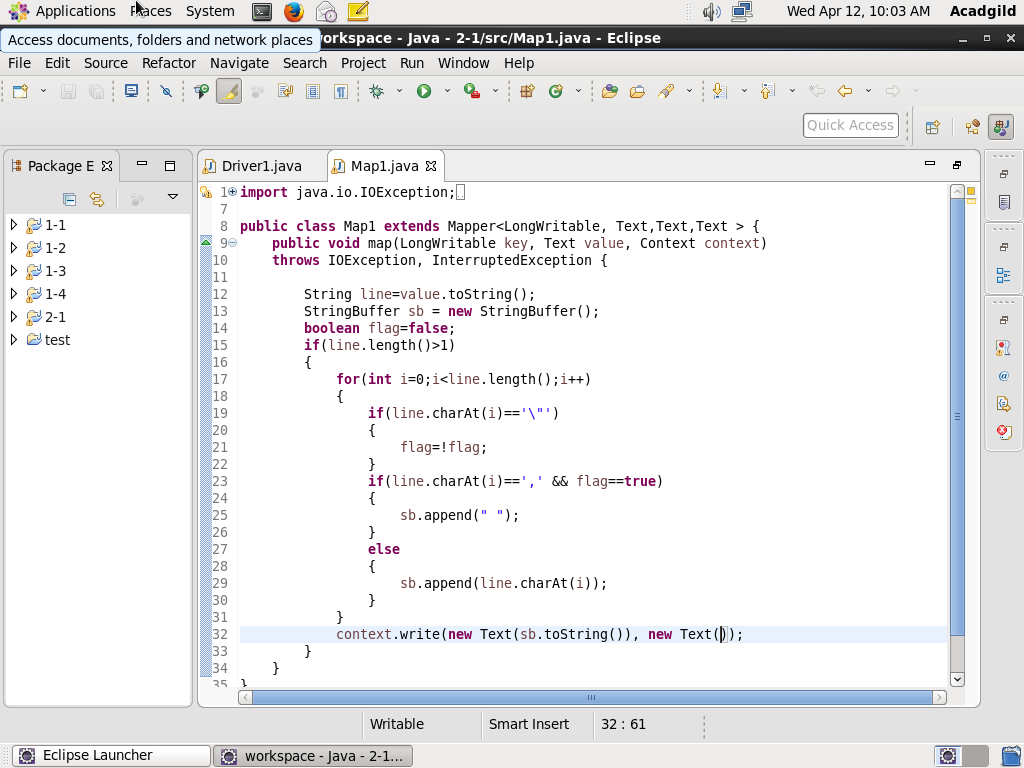
**Q1-Write a mapreduce program to remove commas present inside the double quotes.**

**PROGRAM:**

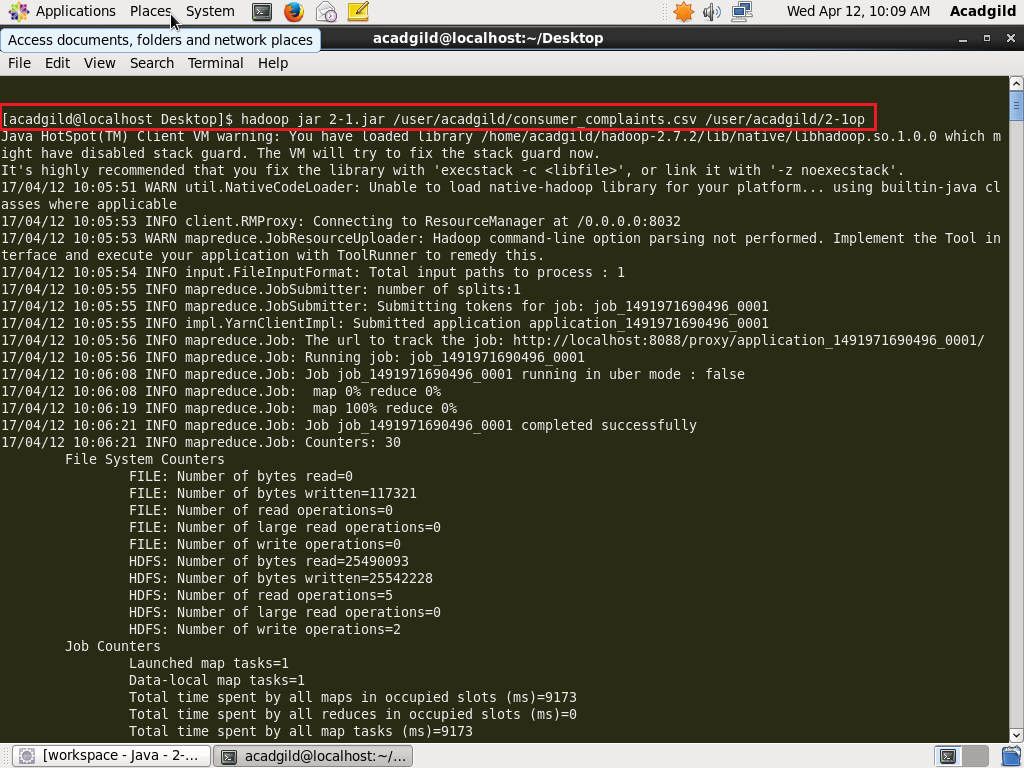
**DRIVER CLASS:**

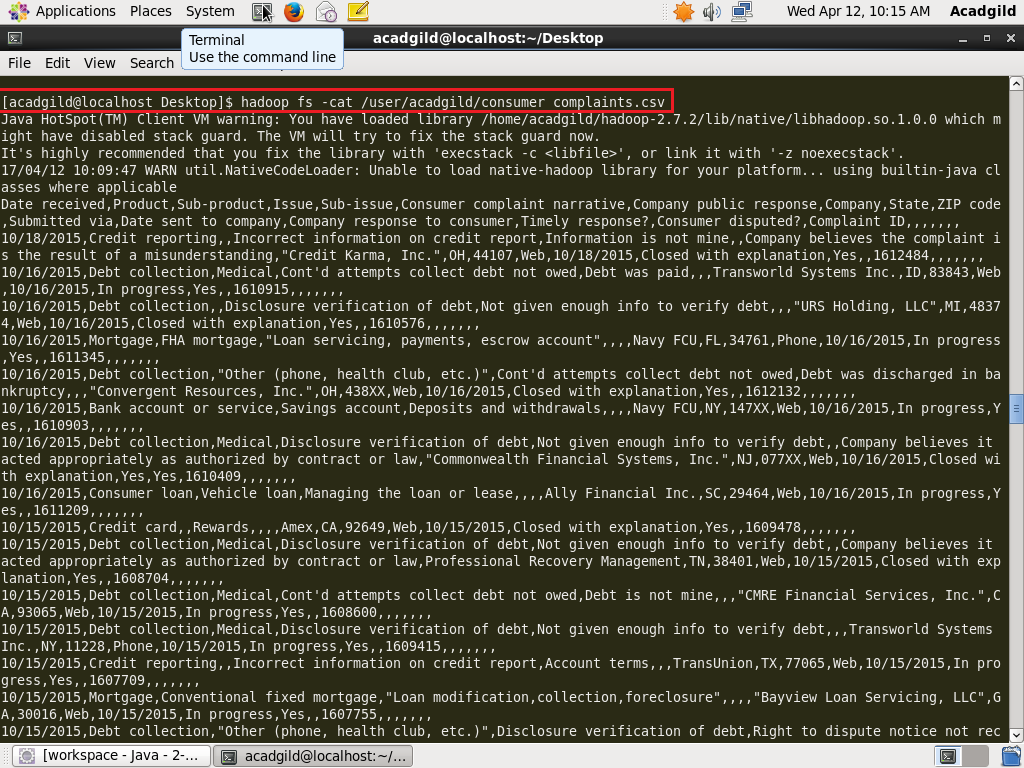
****

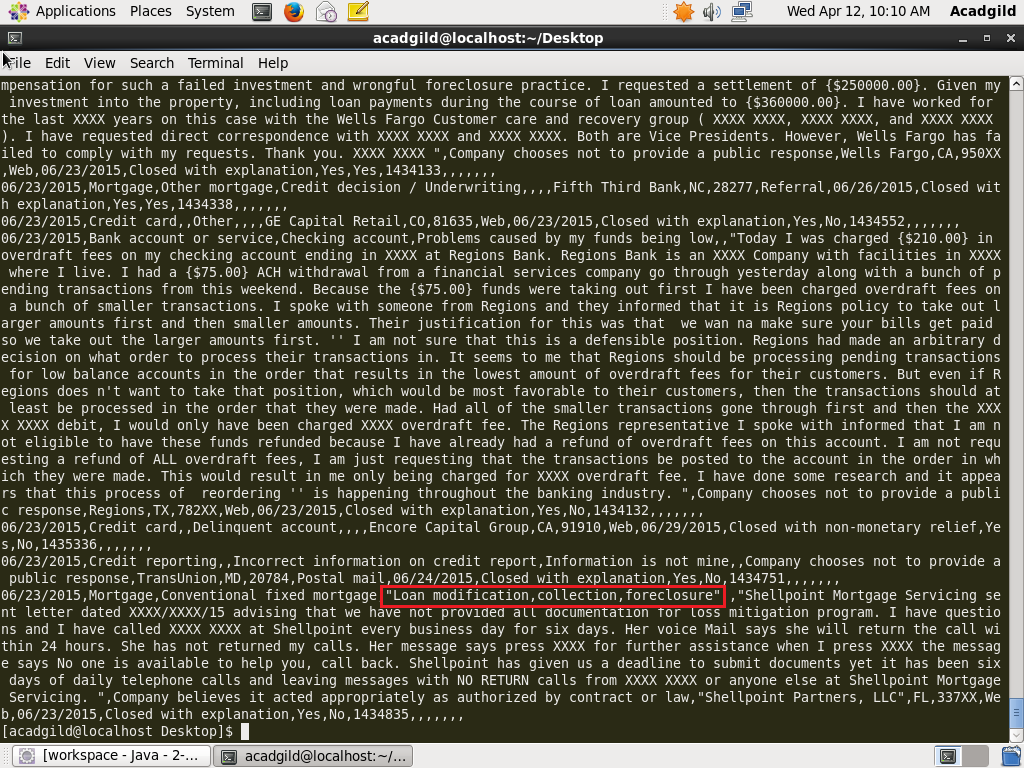
**MAPPER CLASS:**

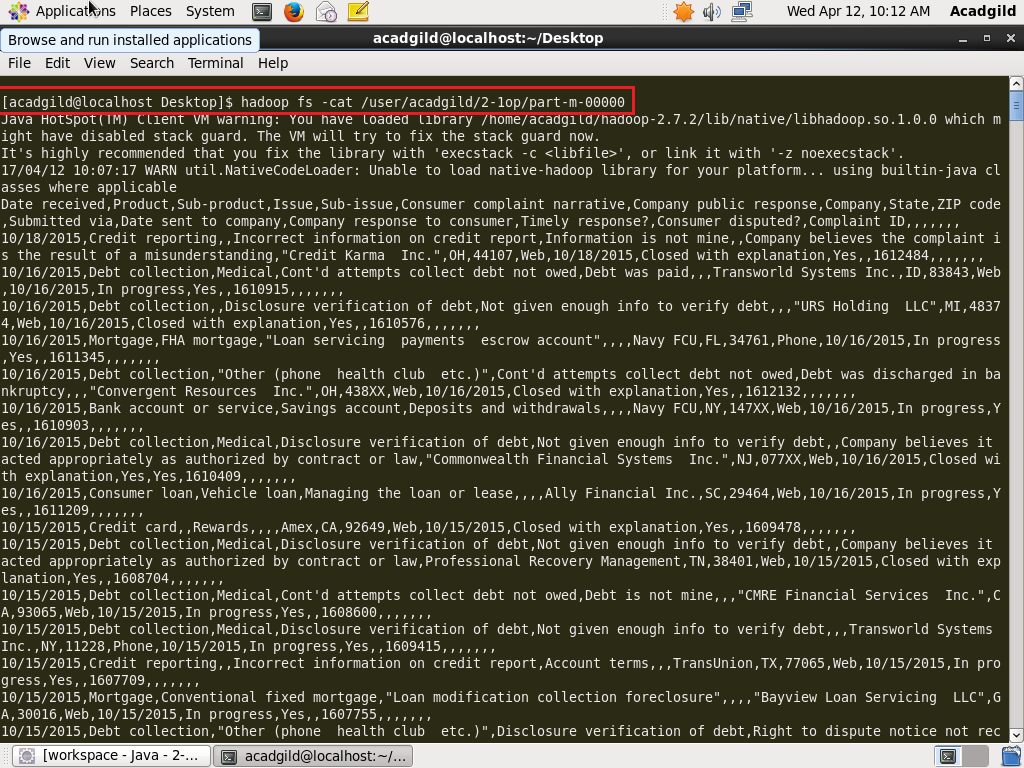
****

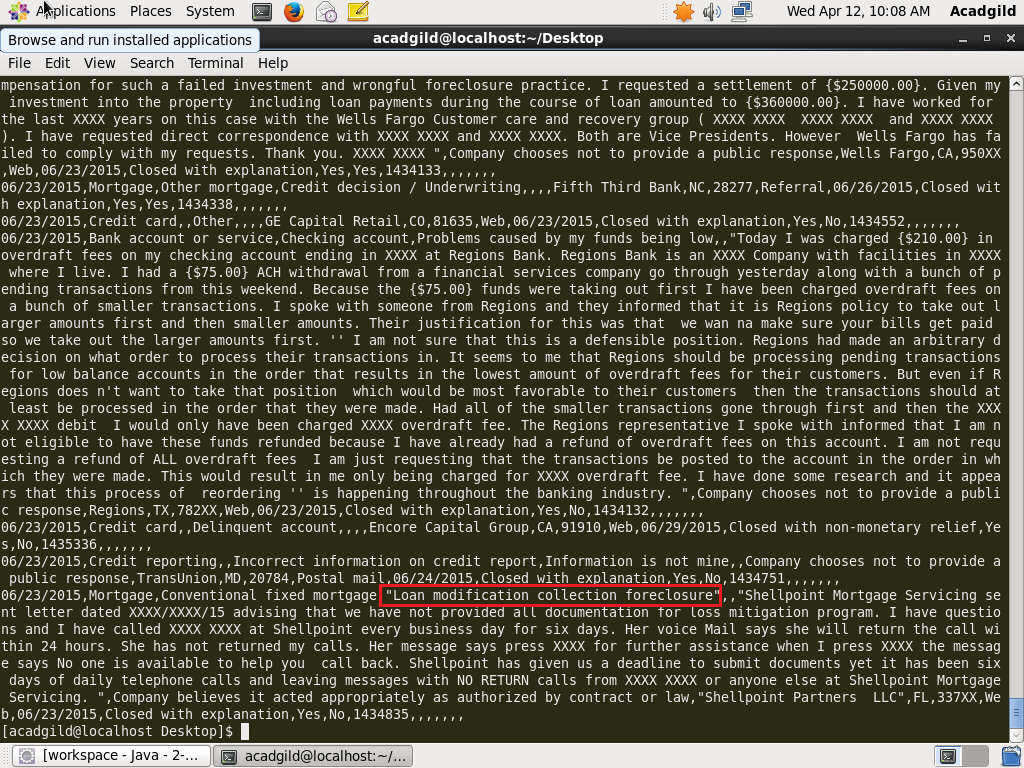
**EXECUTION COMMAND:**

****

****

****

****

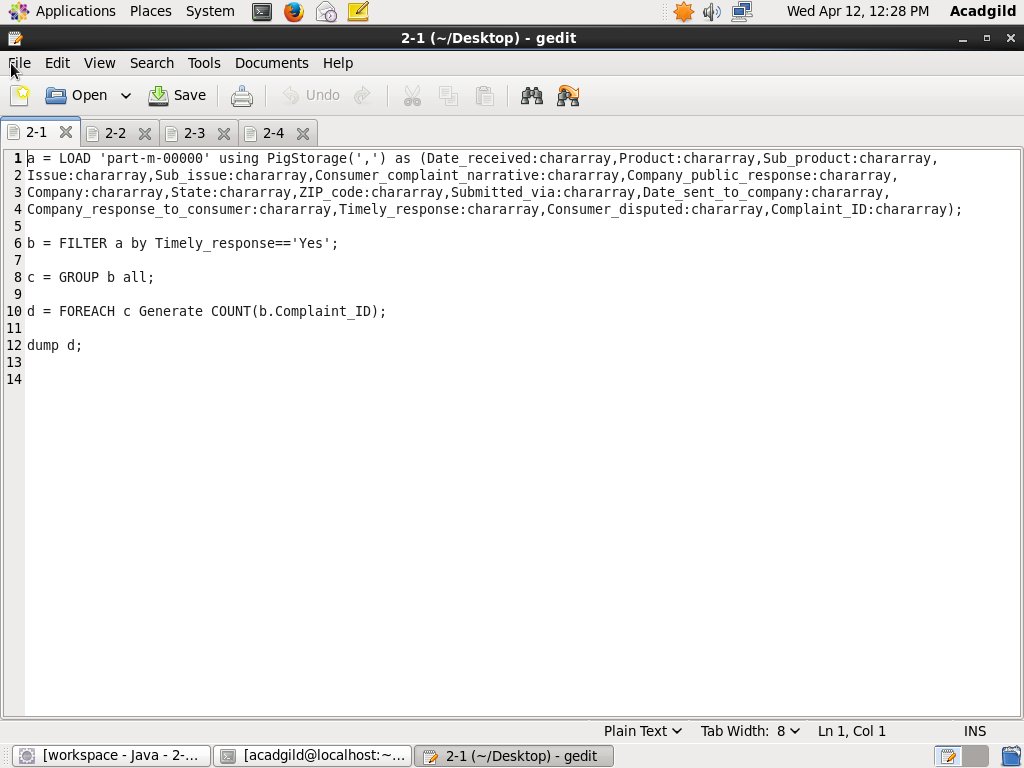
****

**The commas within double quotes are eliminated.**

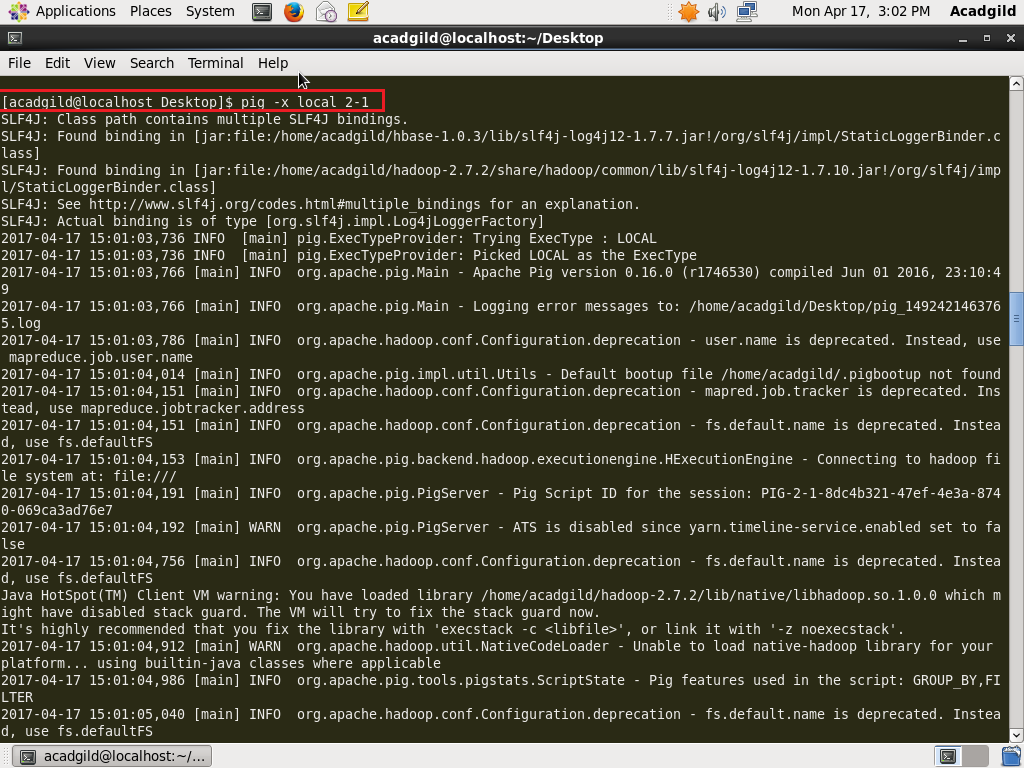
**SECTION-2:**

**Q1- Write a pig script to find no of complaints which got timely response.**

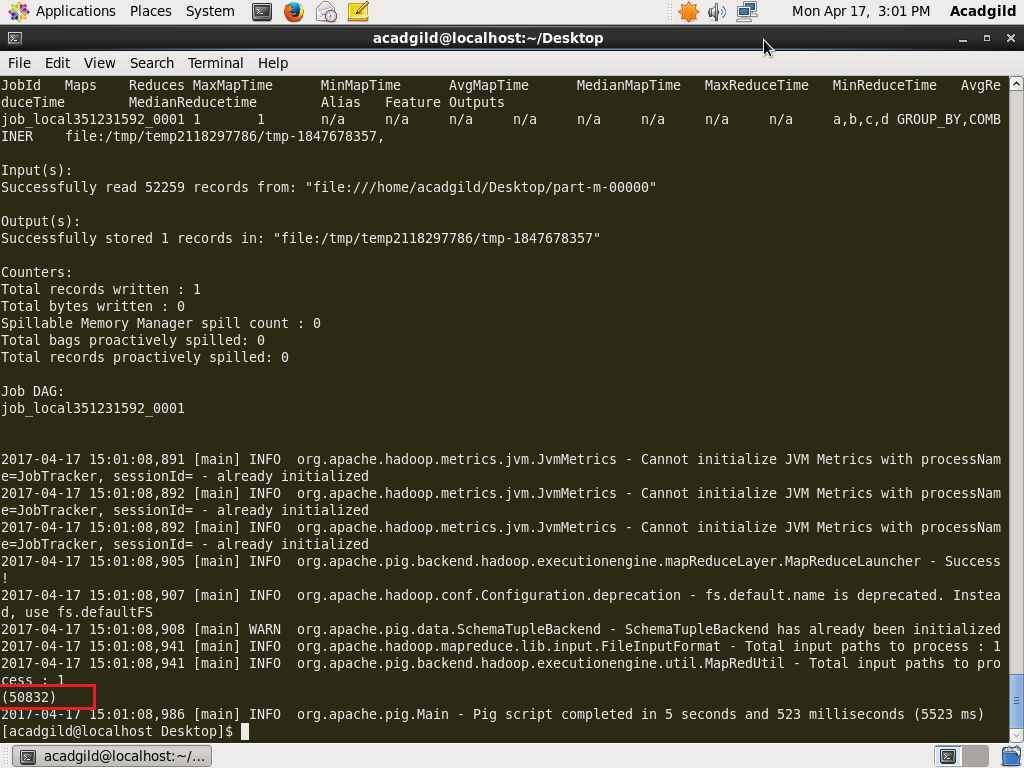
**PROGRAM:**

****

**EXECUTION COMMAND:**

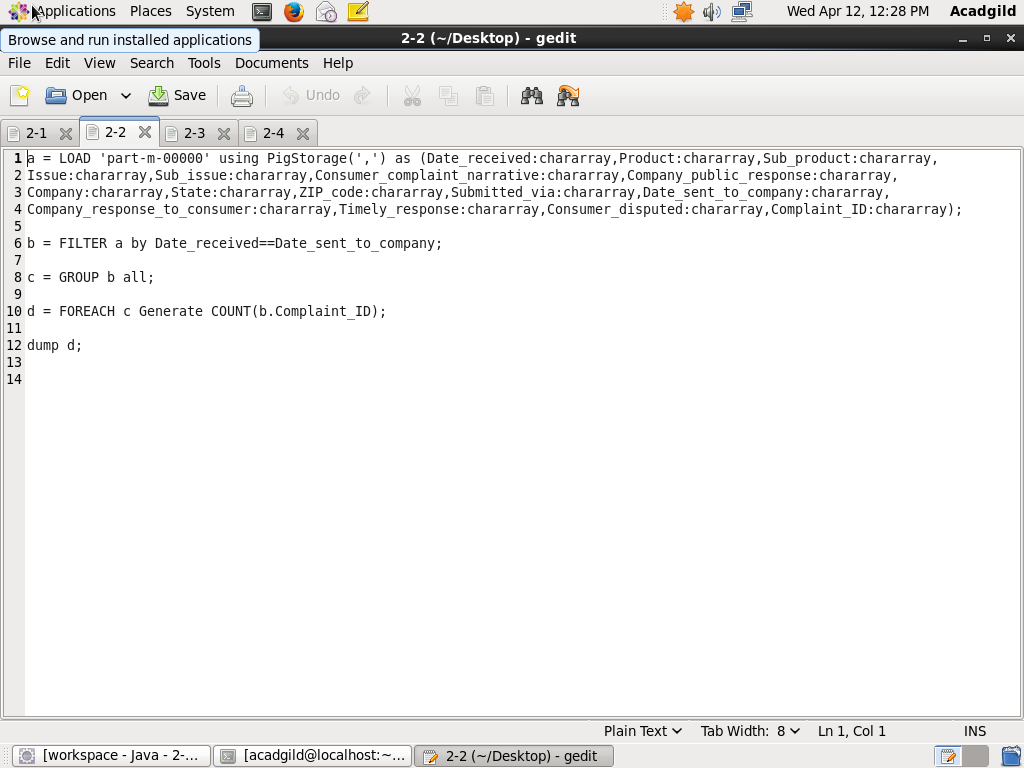
****

**OUTPUT:**

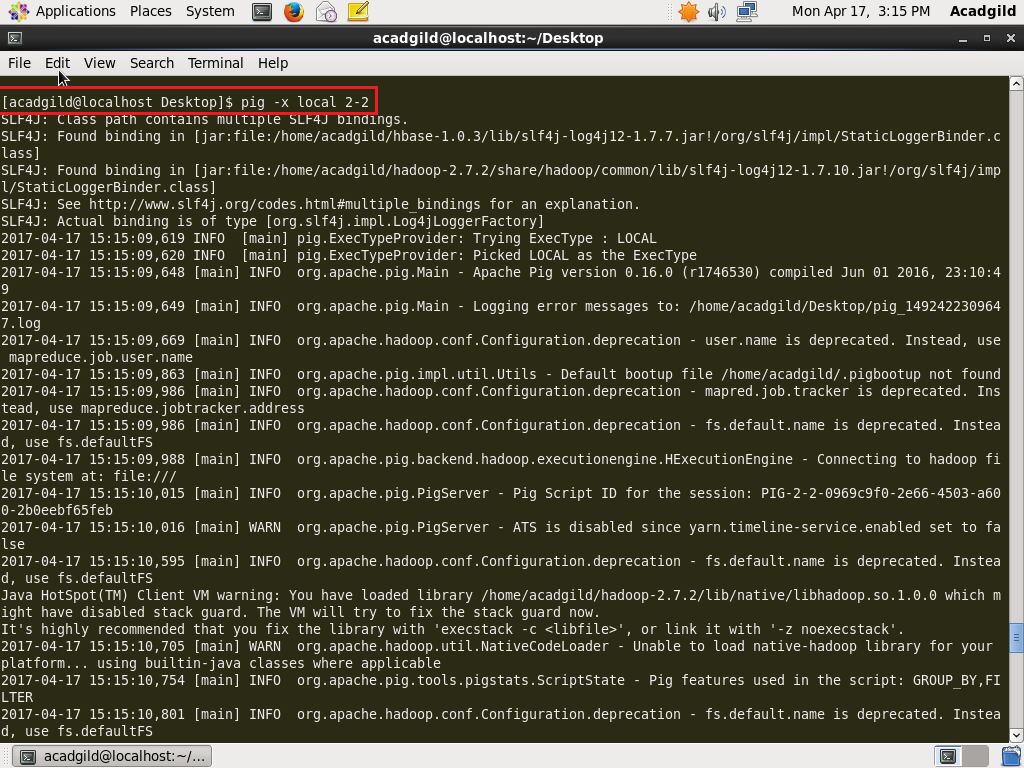
****

**Q2- Write a pig script to find no of complaints where consumer forum forwarded the complaint same day they received to respective company.**

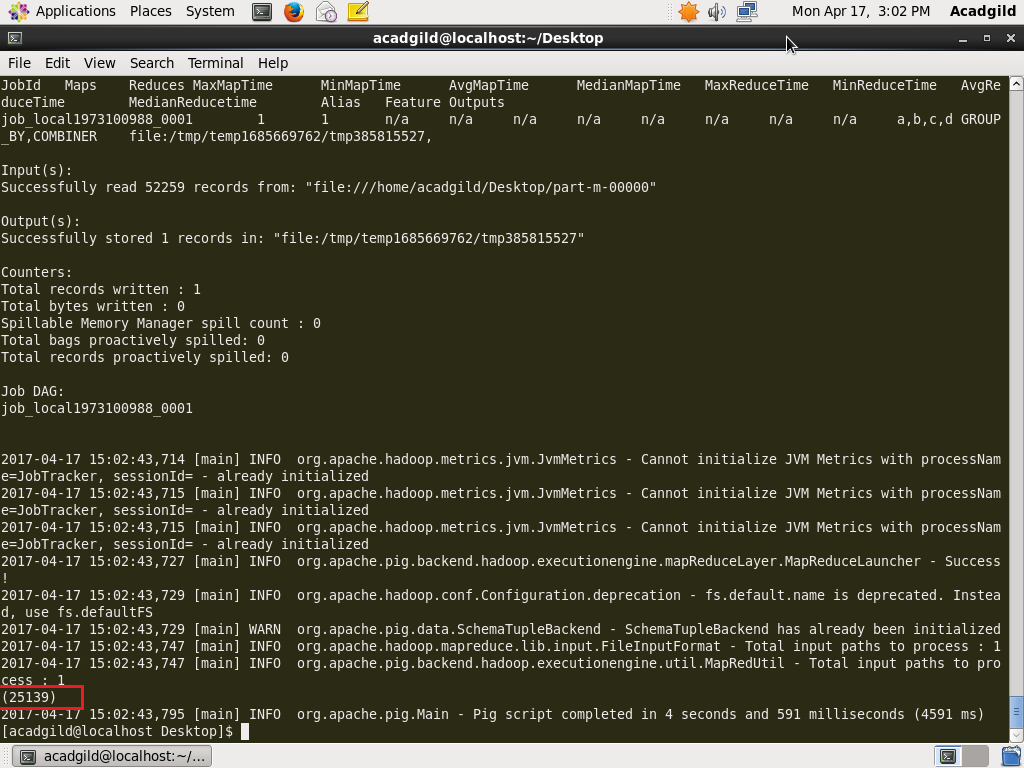
**PROGRAM:**

****

**EXECUTION COMMAND:**

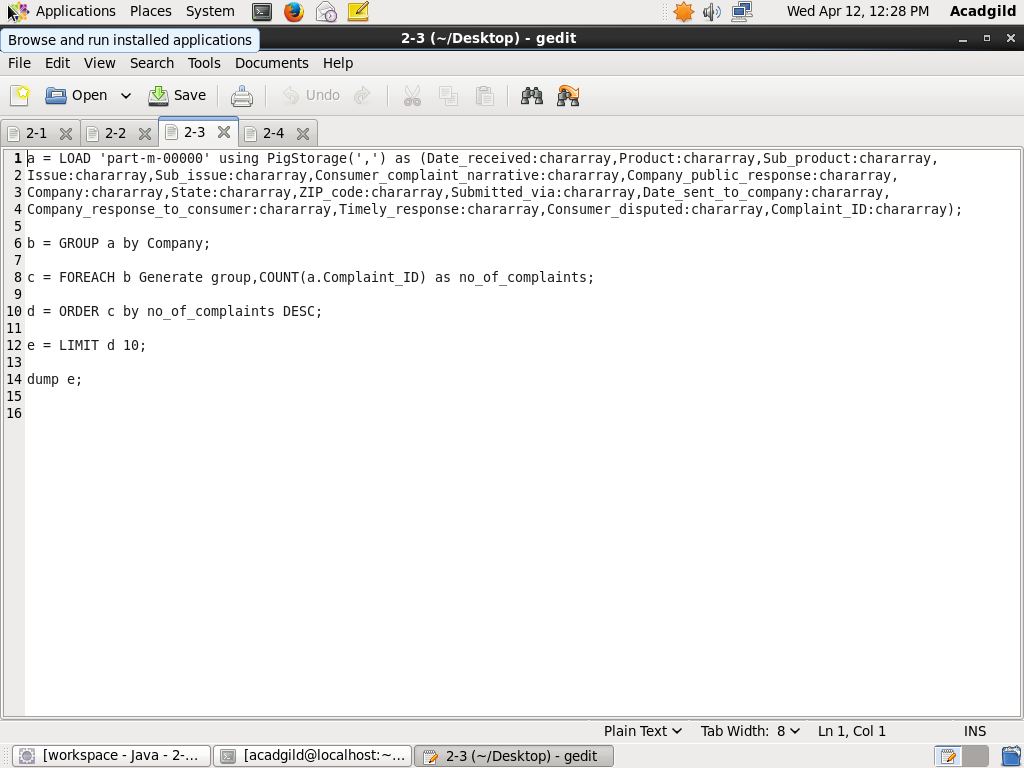
****

**OUTPUT:**

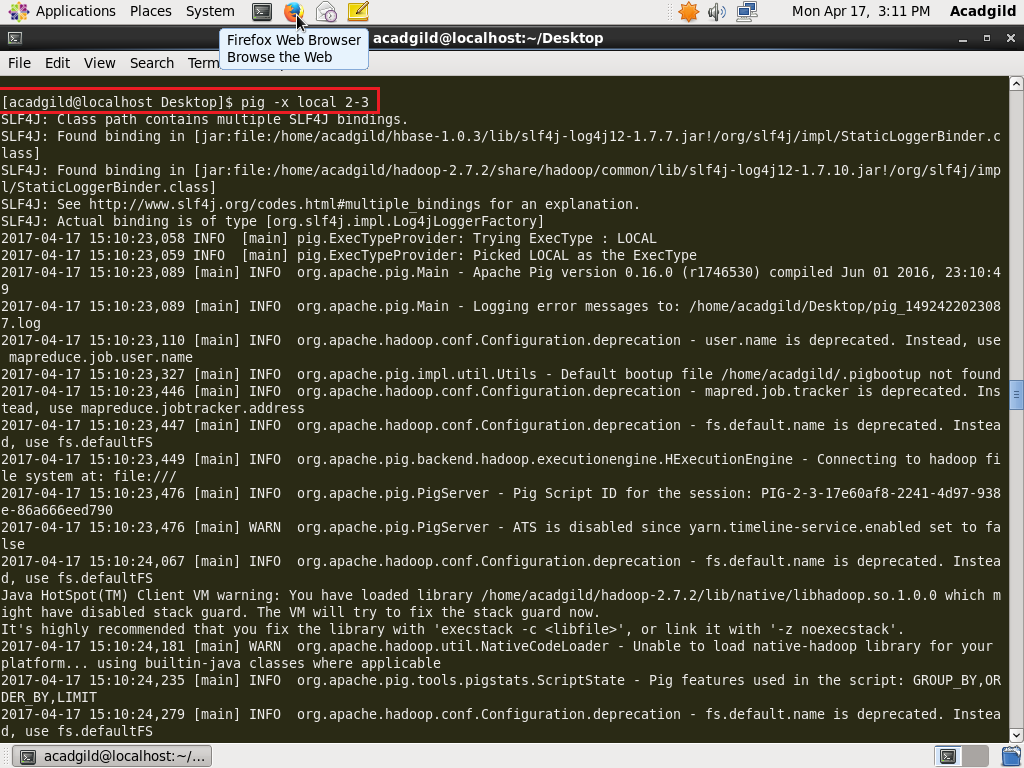
****

**Q3- Write a pig script to find list of companies topping in complaint chart(companies with maximum number of complaints).**

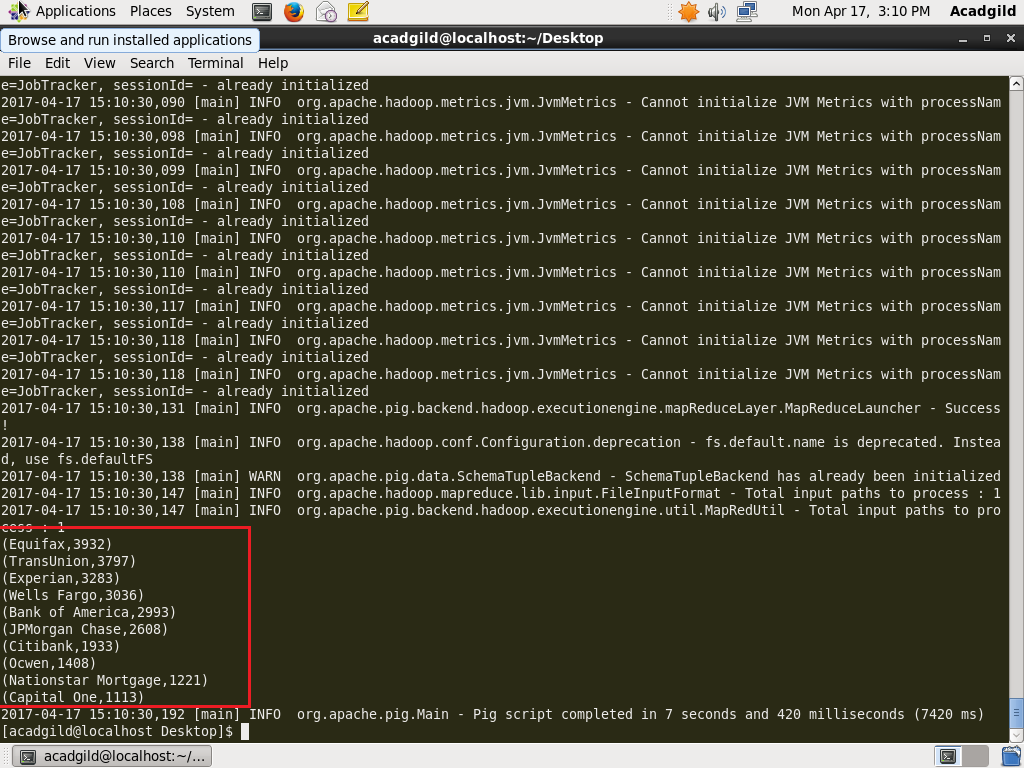
**PROGRAM:**

****

**EXECUTION COMMAND:**

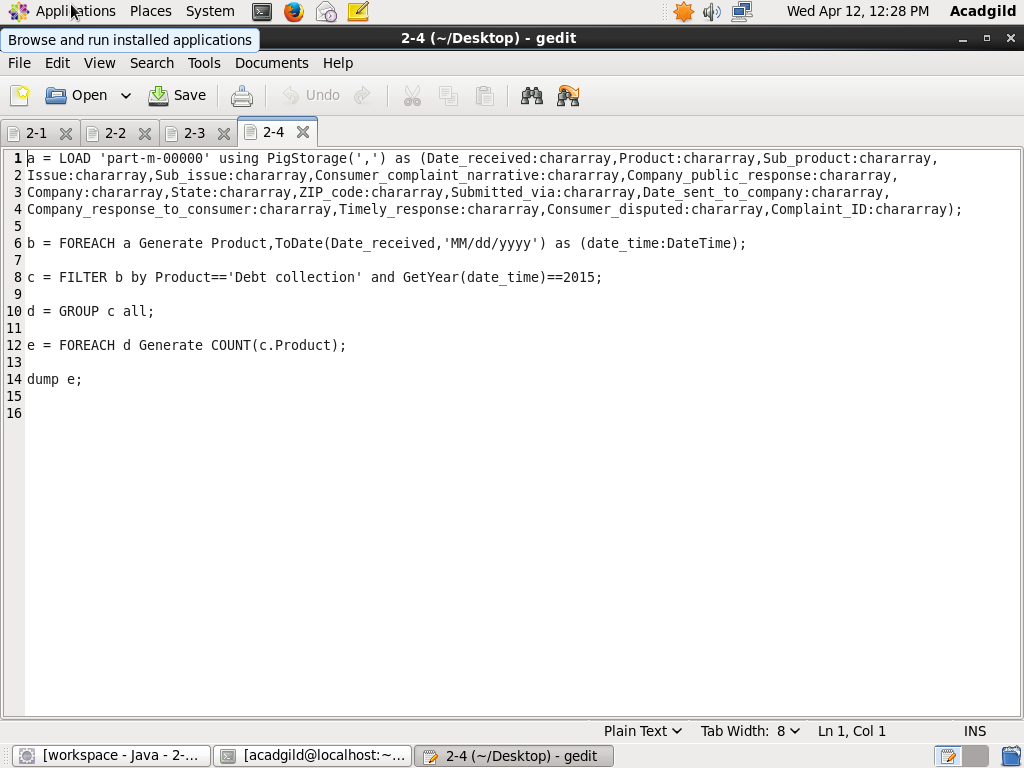
****

**OUTPUT:**

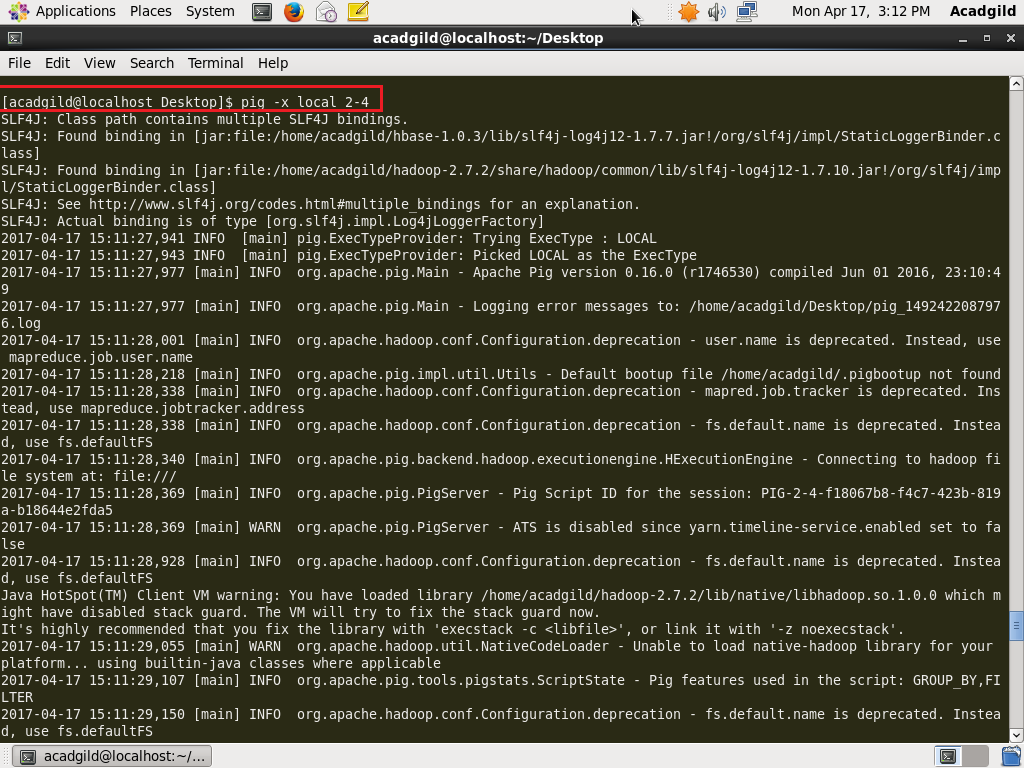
****

**Q4- Write a pig script to find no of complaints filed with product type has "Debt collection" for the year 2015.**

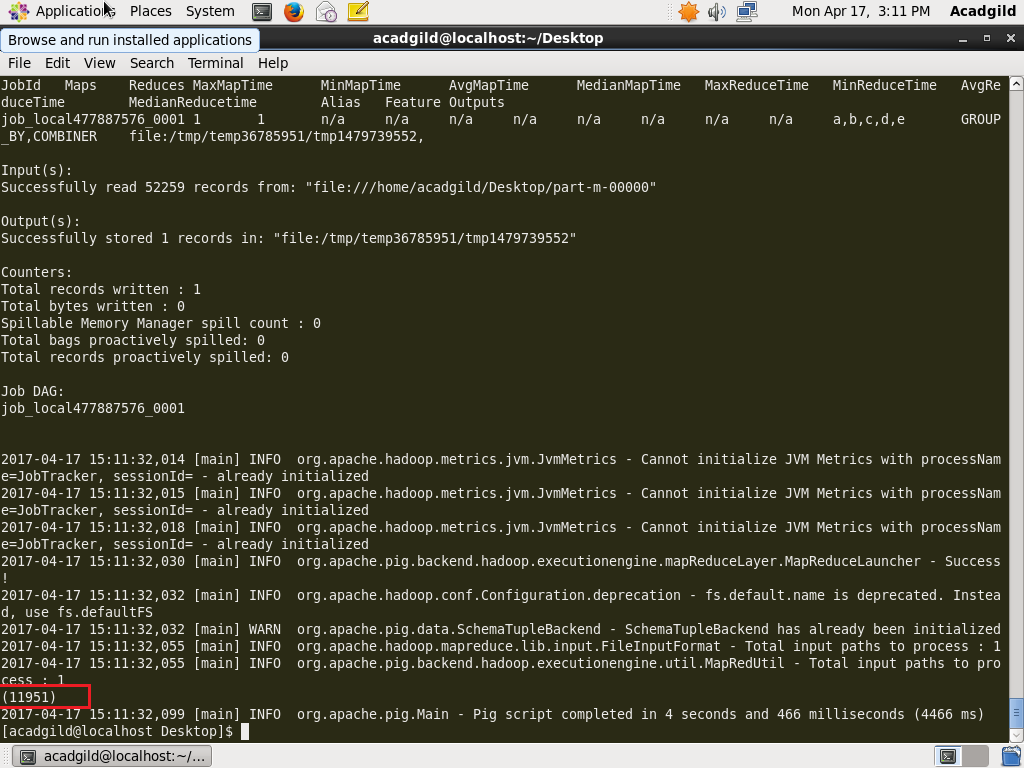
**PROGRAM:**

****

**EXECUTION COMMAND:**

****

**OUTPUT:**

****