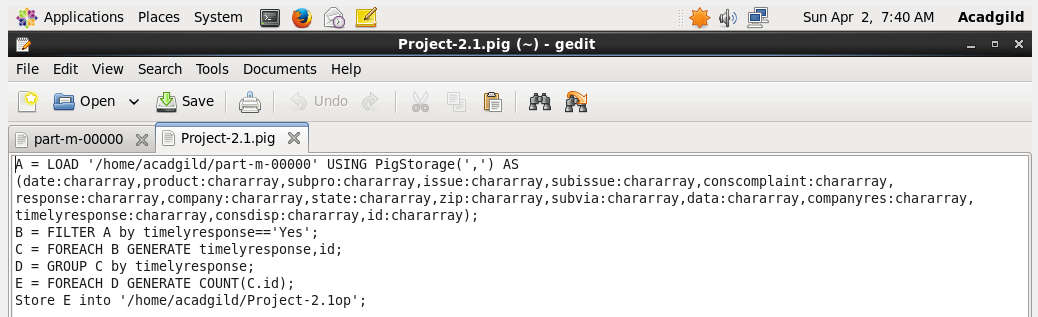
**PROJECT-2**

**USA CONSUMER DATA ANALYSIS**

Project Question-1

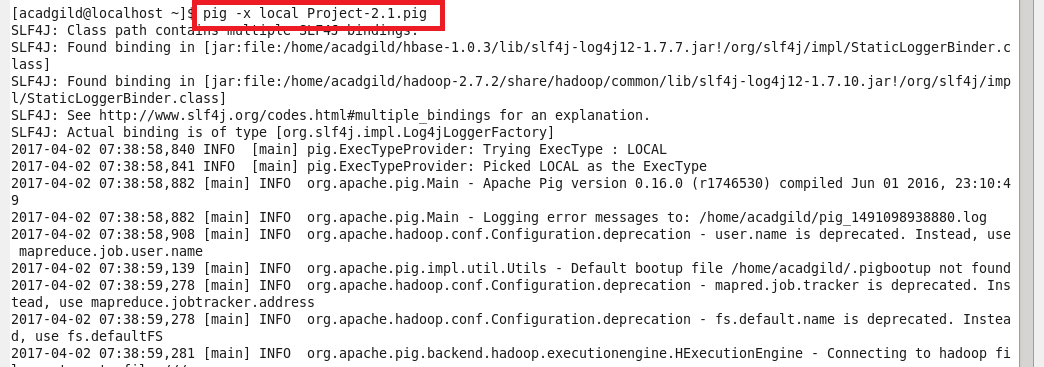
1. Write a pig script to find no of complaints which got timely response

**Pig Script:**

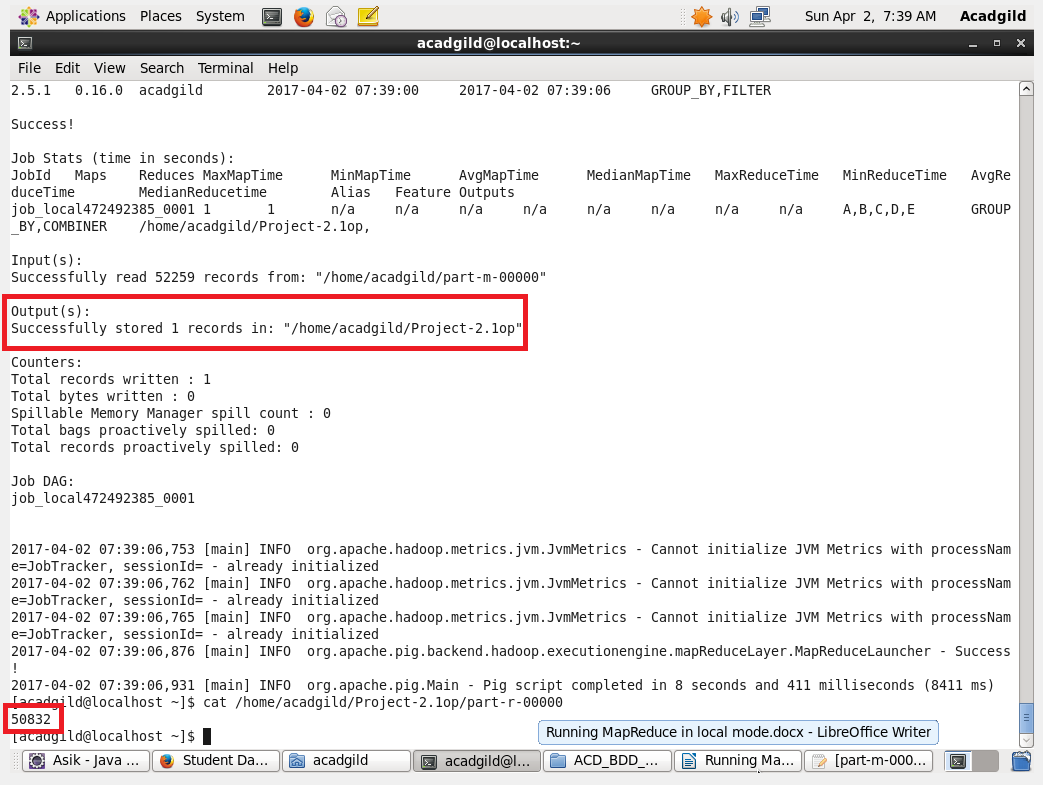


In this script, First the comma removed data is loaded and then it is filtered for the timely response==Yes. Then grouping the data by Timely response and then generating the count of the id. This will give the number of complaints which got timely response.

**Executing the Pig Script in Local Mode:**



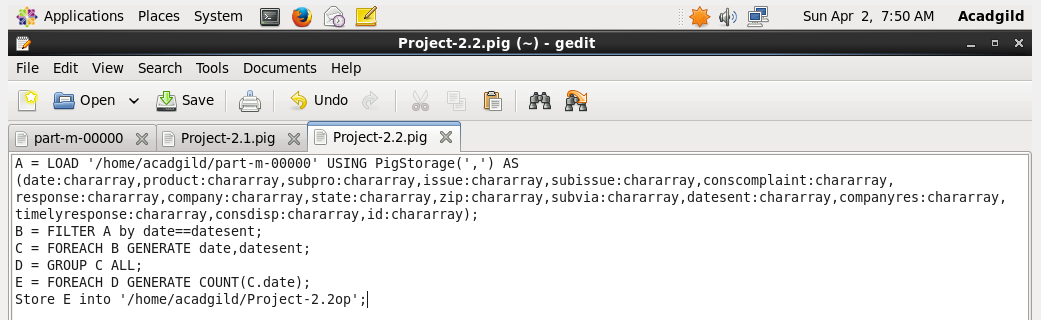
**Output:**



The number of complaints which got timely response are 50832

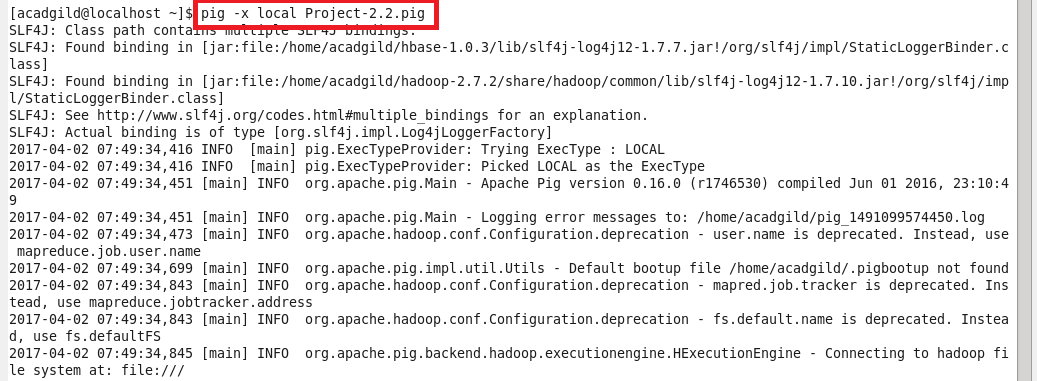
2. Write a pig script to find no of complaints where consumer forum forwarded the complaint same day they received to respective company

**Pig Script:**

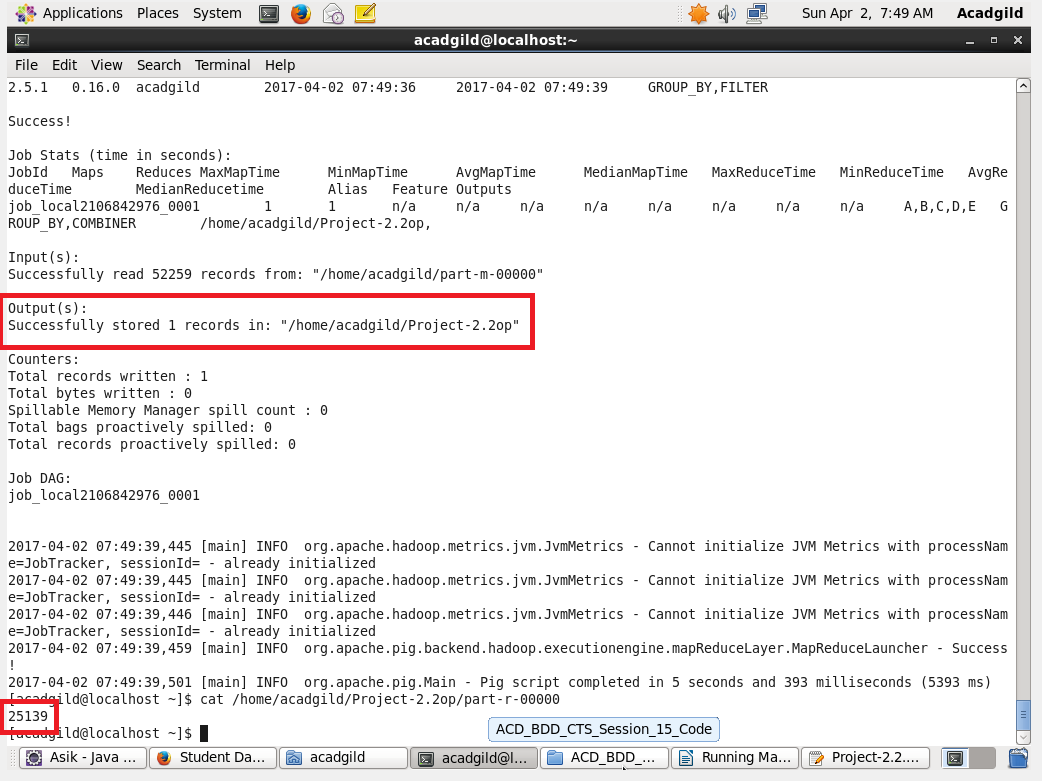


The data is loaded and it is filtered based on this condition date received = date sent. Then the data is grouped and the date is counted and this will give the no of complaints where consumer forum forwarded the complaint same day they received to respective company

**Running the Script in the Local Mode:**

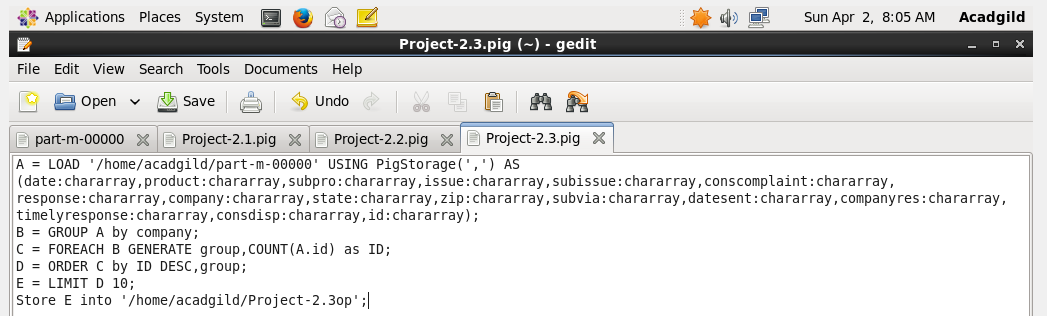


Output:



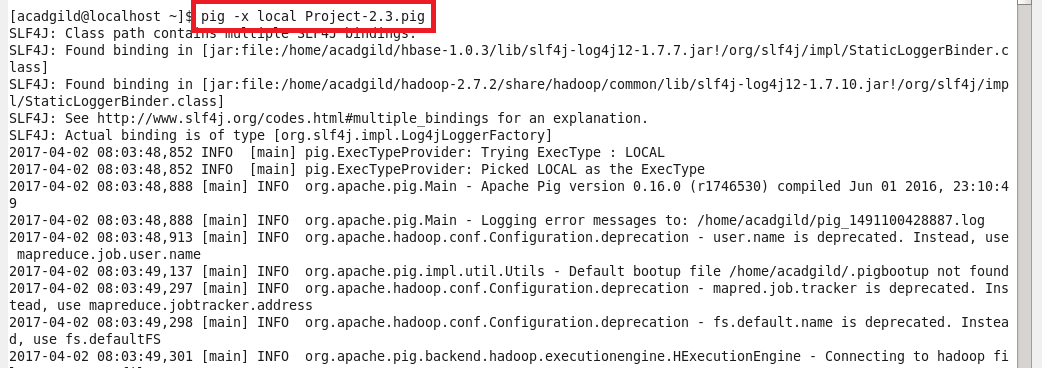
3. Write a pig script to find list of companies topping in complaint chart (companies with maximum number of complaints).

**Pig Script:**

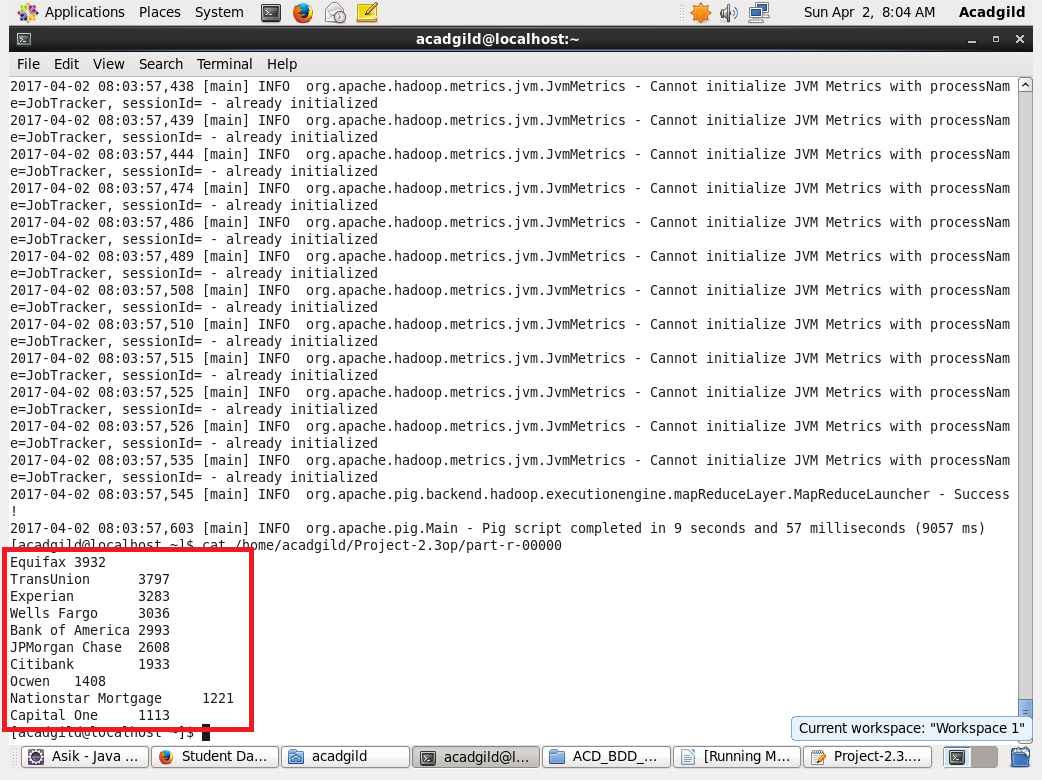


The data is loaded, then it is grouped by the company. And the ID is counted, after that the data is sorted in the descending order this way we can get the maximum no of complaints. Then the data is limited to 10.

**Running the Pig Script:**



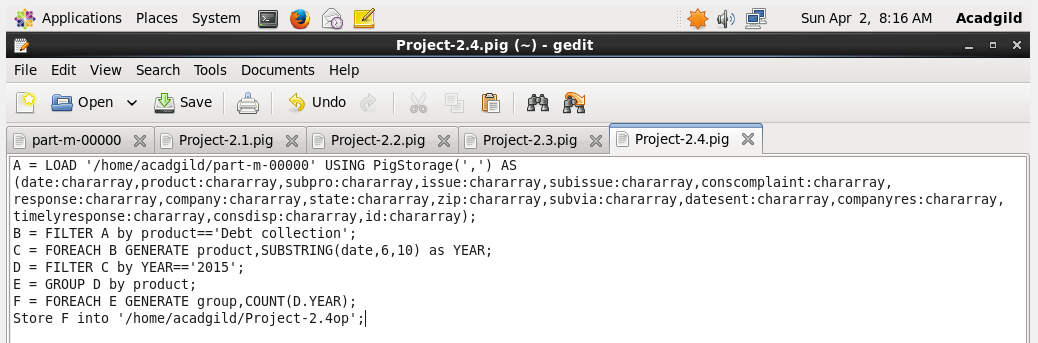
**Output:**



The above screenshot gives the top 10 companies that tops in the complaint chart (companies with maximum number of complaints)

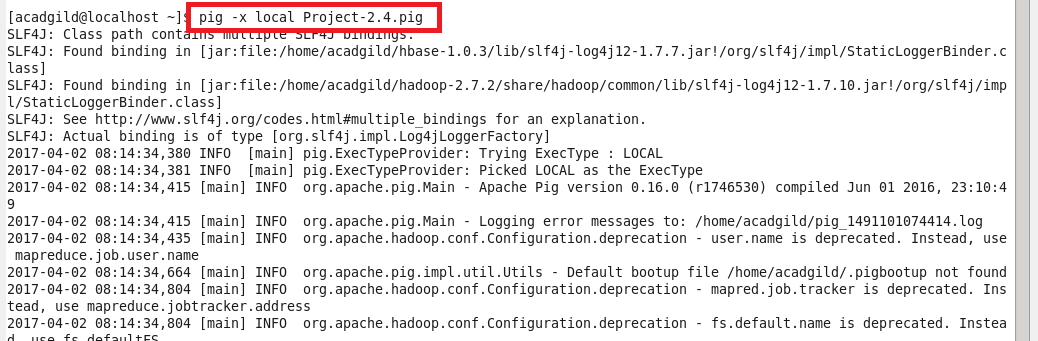
4. Write a pig script to find no of complaints filed with product type has "Debt collection" for the year 2015

**Pig Script:**



In this problem, the data is loaded and filtered based on product==’Debt collection’. Then we have to take the year from the date so that it will be sorted by the year==’2015’. After that the data is grouped by the product and it is counted.

**Running the Pig Script:**



**Output**:

