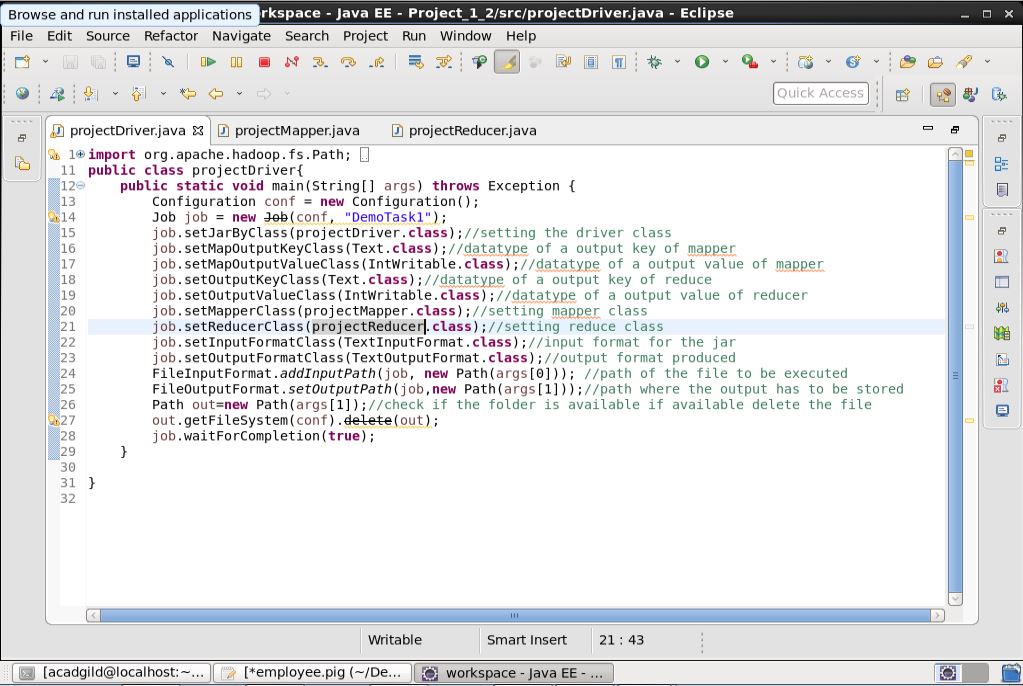
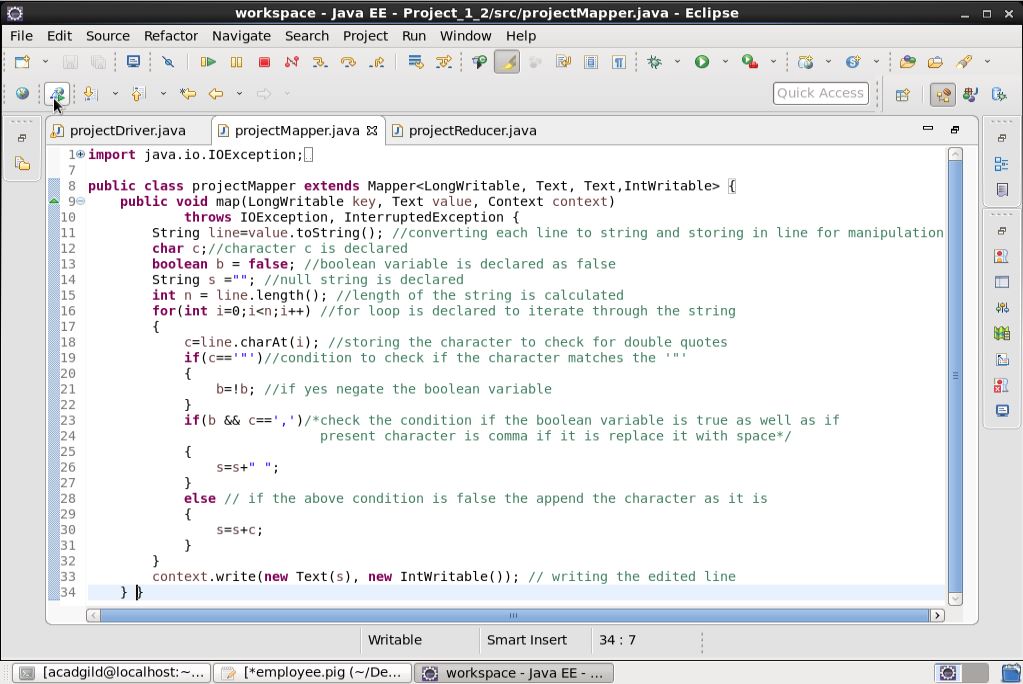
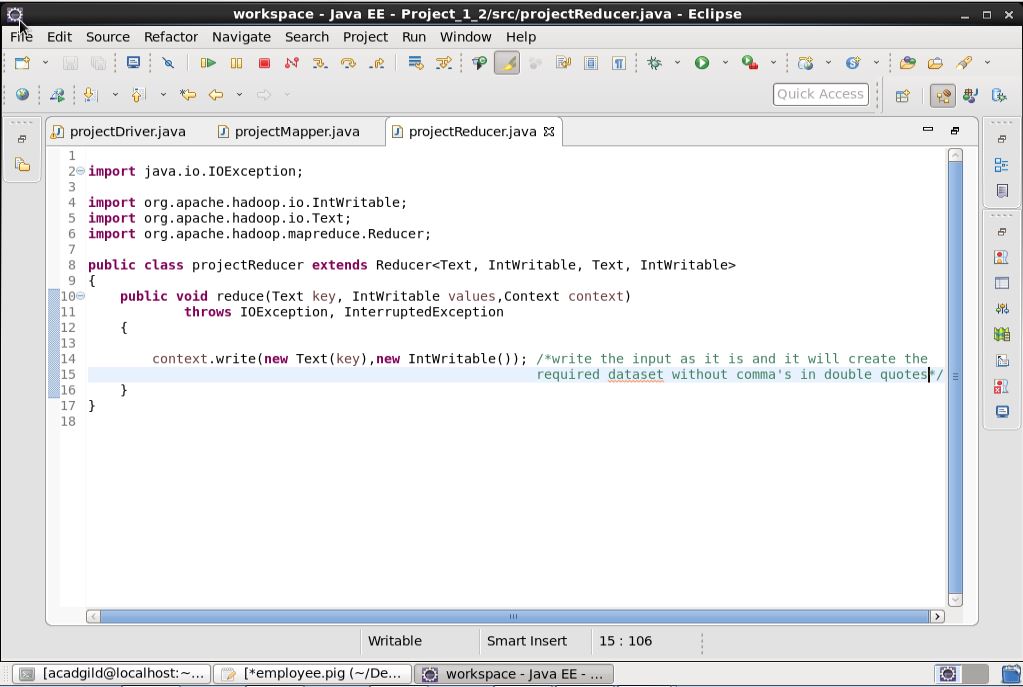
1. Write a mapreduce program to remove commas present inside the double quotes.

We need to delete the commas that are present in the double quote(“) from the given dataset. For this task we have written a mapreduce code. The driver class is used to set the configuration and the input output file formats and the classes of mapper and reducer and many different configurations.

The mapper class is implemented to delete the commas that are present in the double quotes(“), for that we have implemented the logic using the Boolean variable and converting the string to a character. We will check for the double quote occurrence after the first occurrence the Boolean variable is toggled and the comma occurring after the first occurrence will be replaced by white space and once second double quote is encountered the commas will not be deleted as we will toggle the Boolean variable. This is how we will be creating the dataset for the pig operations. The dataset generated after the deletion of the unwanted commas we will feed this dataset as an input to the pig scripts.







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

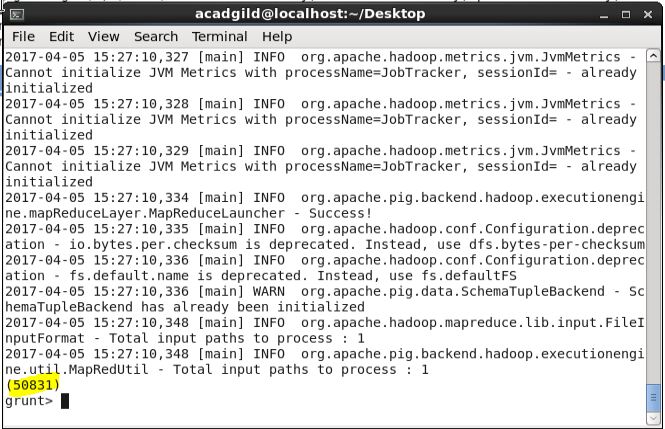
Here in this question we need to find the complaints those got timely response. To know we need to check if the column is stating ‘yes’ or ‘no’. If it is yes then the complaint got timely response.

The pig script for the required question is as follows:



In this pig script we have loaded the data that we created using the mapreduce as the input dataset. As we need to split by comma. Now we will check if the complaint got the timely response by checking the condition for the column timely response if it is yes. To check used use ‘filter ‘ command and the dataset is filtered accordingly. Then to get the count we first generated the group by timely response. Then the count is generated and we can get the count using the dump command on variable ‘d’.

The answer for this question is 50831.



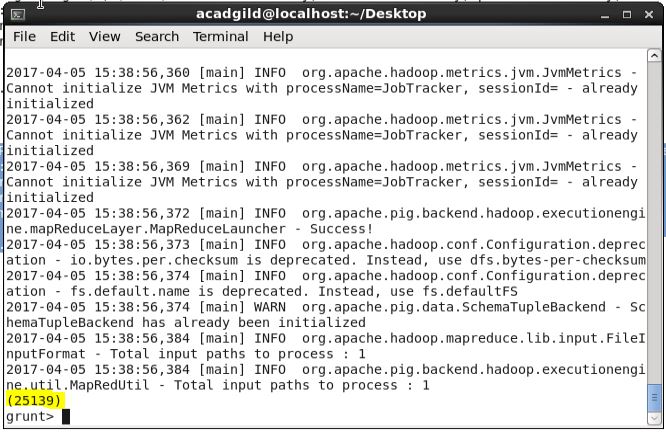
2. Write a pig script to find no of complaints where consumer forum forwarded the complaint same day they received to respective company

In this pig script we will be comparing the date on which the complaint received and the date it was forwarded to the company.

For this firstly we loaded the dataset that we got from the mapreduce. Then the dataset is filtered by comparing the two dates i.e. the date the complaint is received and the date it was forwarded to the company. After the filter operation the dataset generated are of the same dates. Now we will group the filtered data by all and lastly the count is generated using the count function.



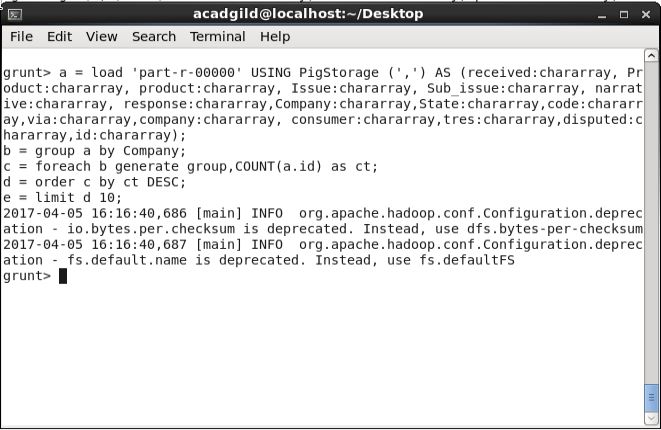
The dump is used on the variable ‘d’ and the count is displayed as:



So the output is 25139.

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

To find the maximum number of complaints we need to find the count of the complaints of each company. For that we have loaded the required dataset and then we grouped it using the Company name. and then the count is generated for each count with the company name preceding. Then the generated name and count is ordered in descending order based on the count of the complaints. And to find the top companies we have considered as top 10 companies and is limited to 10 by using the limit command.



The top 10 companies are as follows :



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

Here to find the number of complaints under the product type as Debt collection we filtered product using product type as Debt Collection.

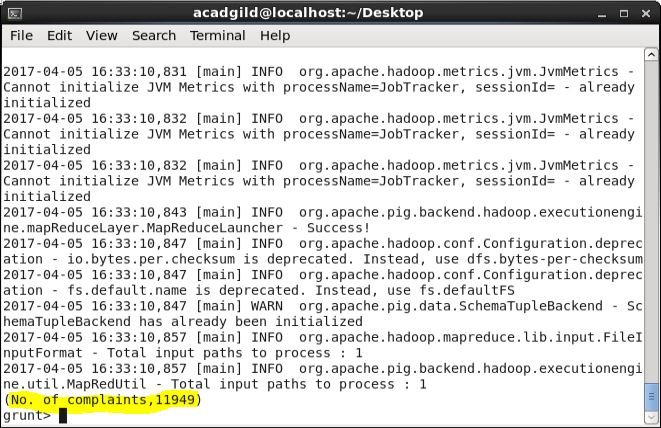
Firstly we loaded the dataset then we have generated the id, product type and converted received to the date format.

Then again we generate the id, product type and the year from the above generated data, to filter according to the required dataset. Now we will filter the dataset based on the product type as ‘Debt Collection’ and the year as 2015. After the filter operation we have the desired data. Now to find the count we first group by product type and then generate the count to find the total number of complaints under Debt collection in the year 2015.

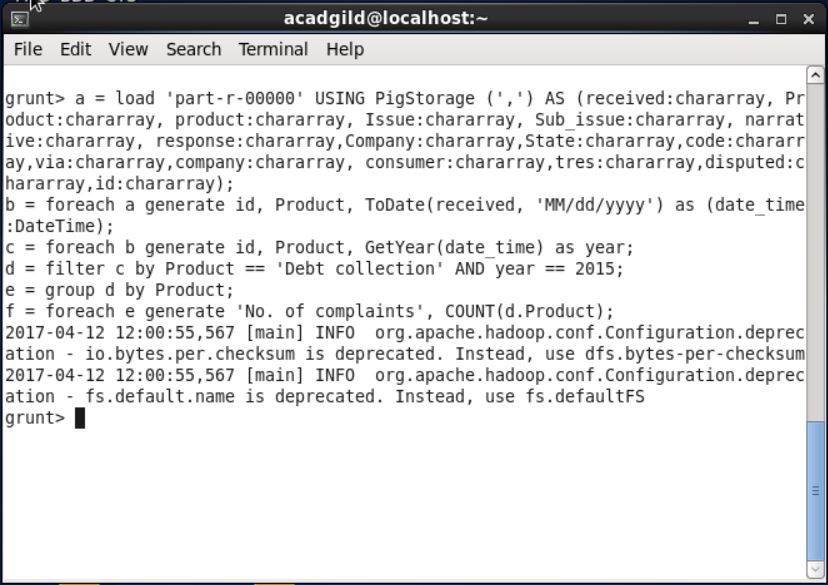
This is the program to count the id’s and the output is 11949. Aslo we have one more code after to this code.

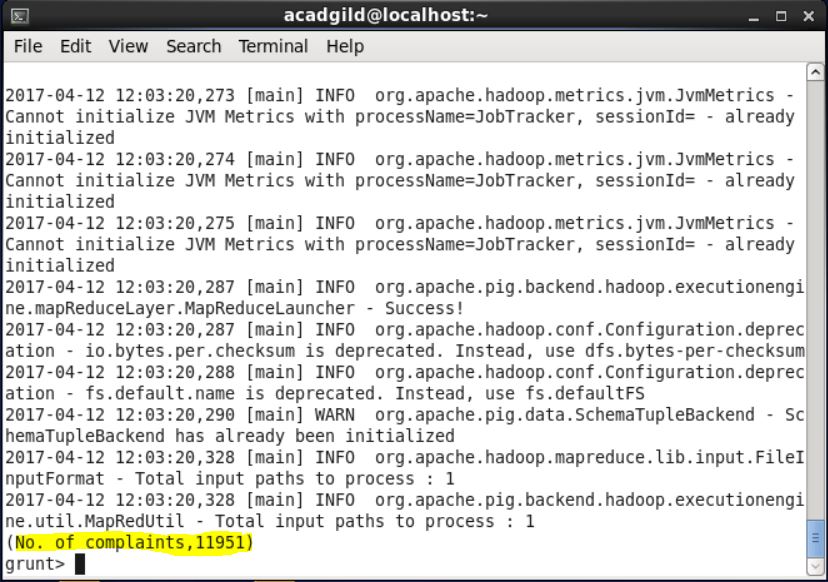


After dumping the variable f we got the number of complaints on the screen as 11949



The code where we count the product and the output is 11951.





The difference of two is because if we consider the id then the closed with explanation is not counted so there is a difference of 2 will occur. If the debt collection is consider then it will consider all the cases that are closed with explanation also. So giving output 1951.