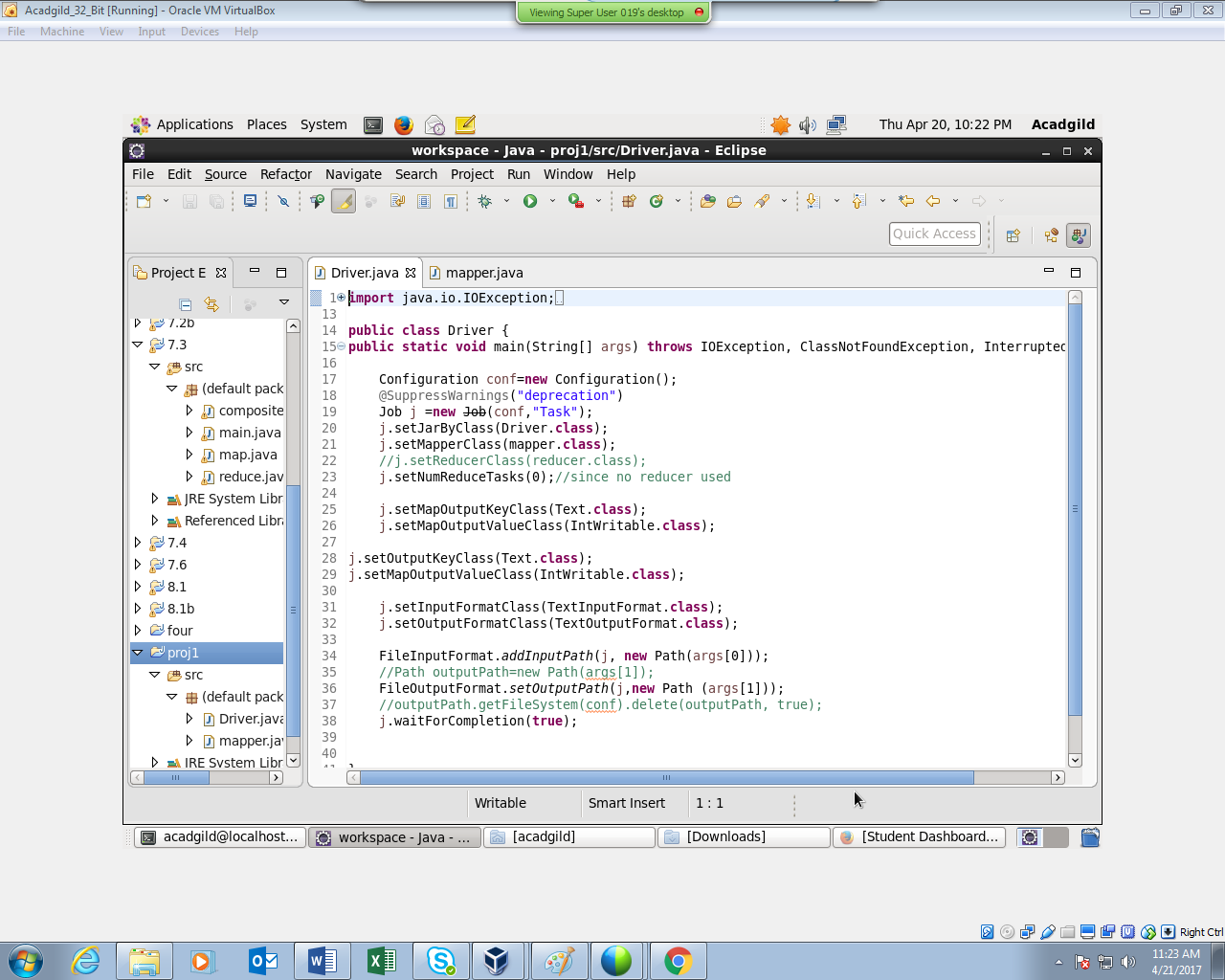
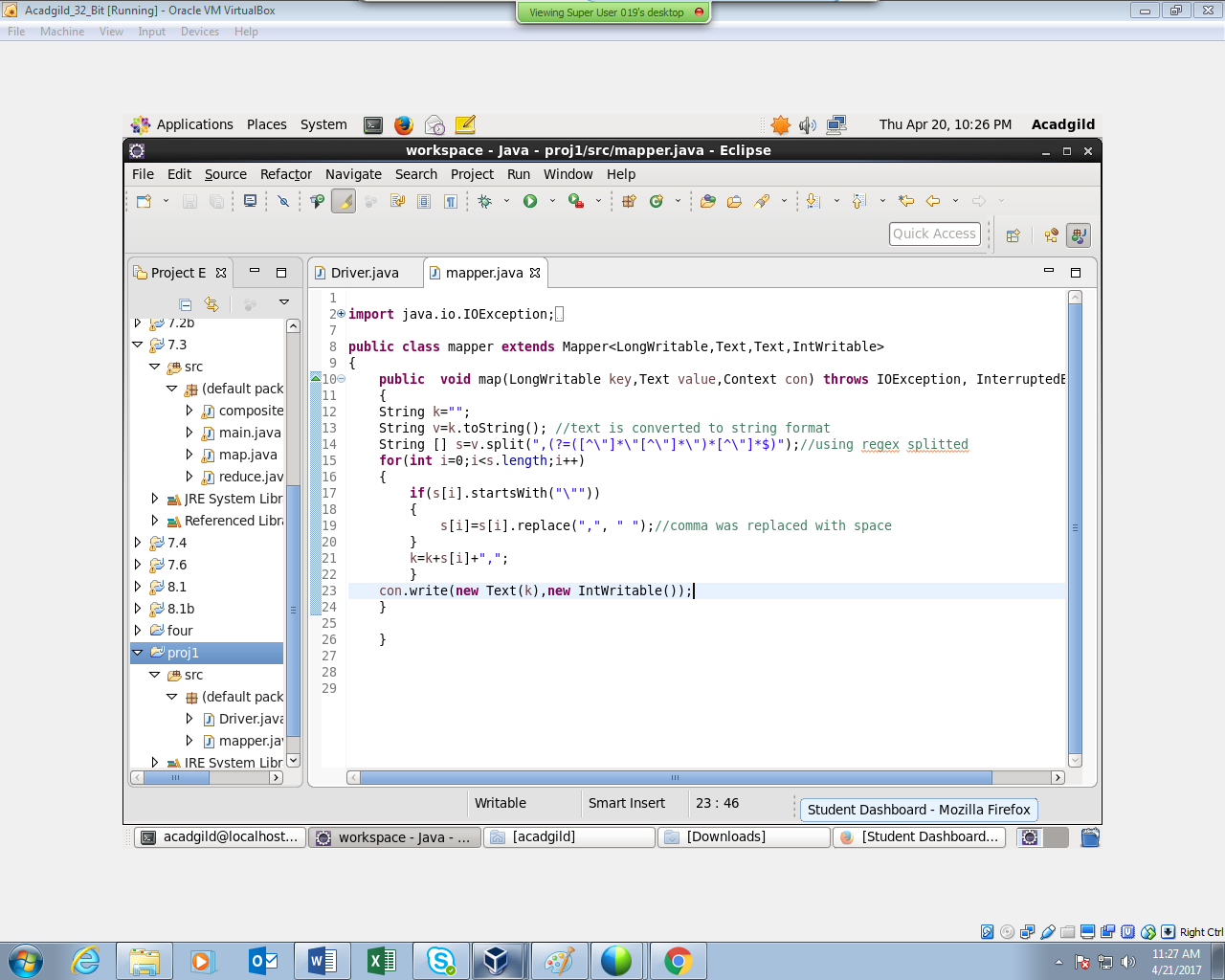
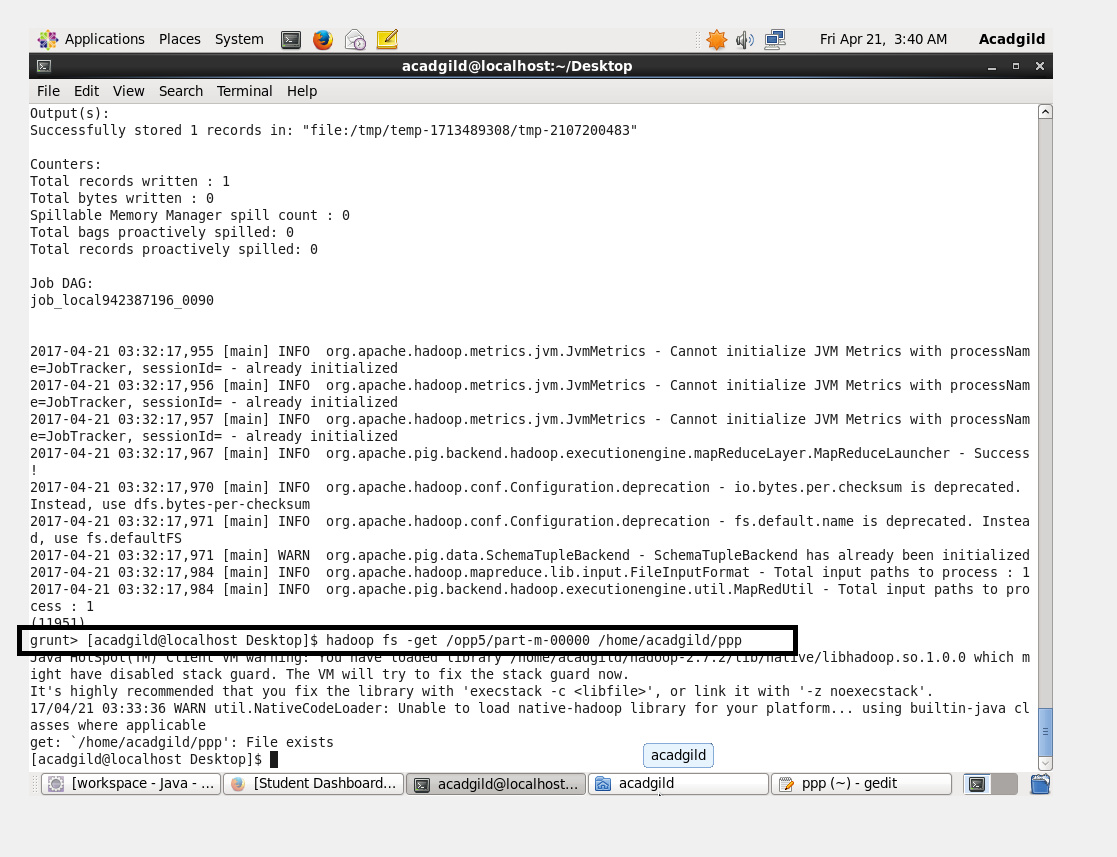
**MAP REDUCE PROGRAM TO REPLACE ALL THE COMMAS INSIDE THE DOUBLE QUOTES WITH SPACE:**

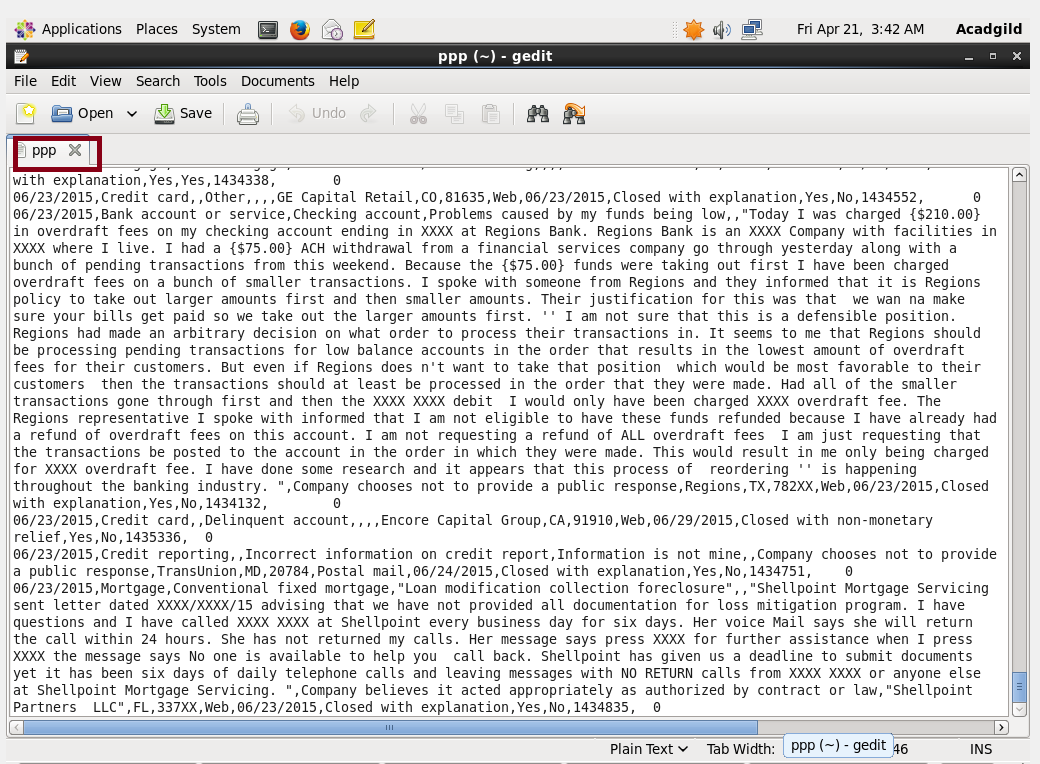
**DRIVER CLASS:**

****

**MAPPER CLASS:**

****

****

****

1. All files will be listed by using hadoop fs –ls /user/acadgild/ppp/.
2. get command – will copy files from hdfs to local.
3. Part-m-00000 is copied to the local in the name of consumer.txt with all the commas inside the double quotes replaced with space.
4. Processed file (ppp) will be used as the Dataset for executing pig commands.
5. **WRITE A PIG SCRIPT TO FIND NO OF COMPLAINTS WHICH GOT TIMELY RESPONSE.**

**LOGIC:**

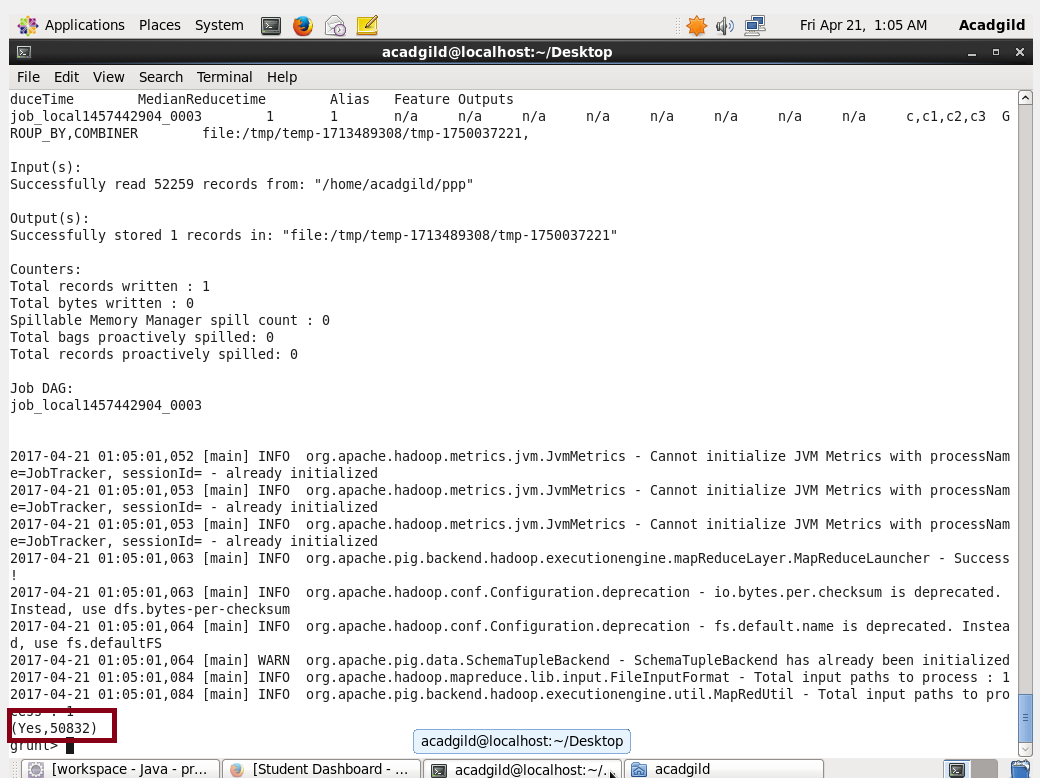
1. Loading the Dataset into the grunt shell using LOAD command.
2. Filtering the Dataset by checking the column 14 is yes.
3. Grouping based on column 14 using Group By command.
4. Counting the number of complaints using COUNT command.

At this stage o/p will be like (Yes, 50832).

1. Generating only the count using GENERATE command by eliminating the yes.



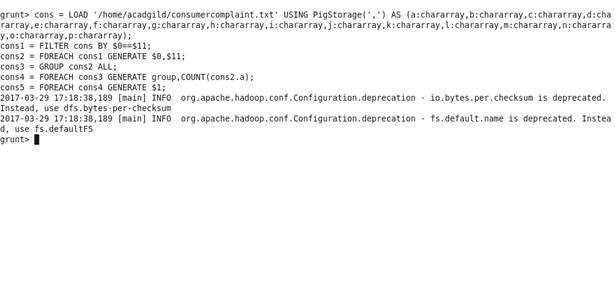
**OUTPUT:**



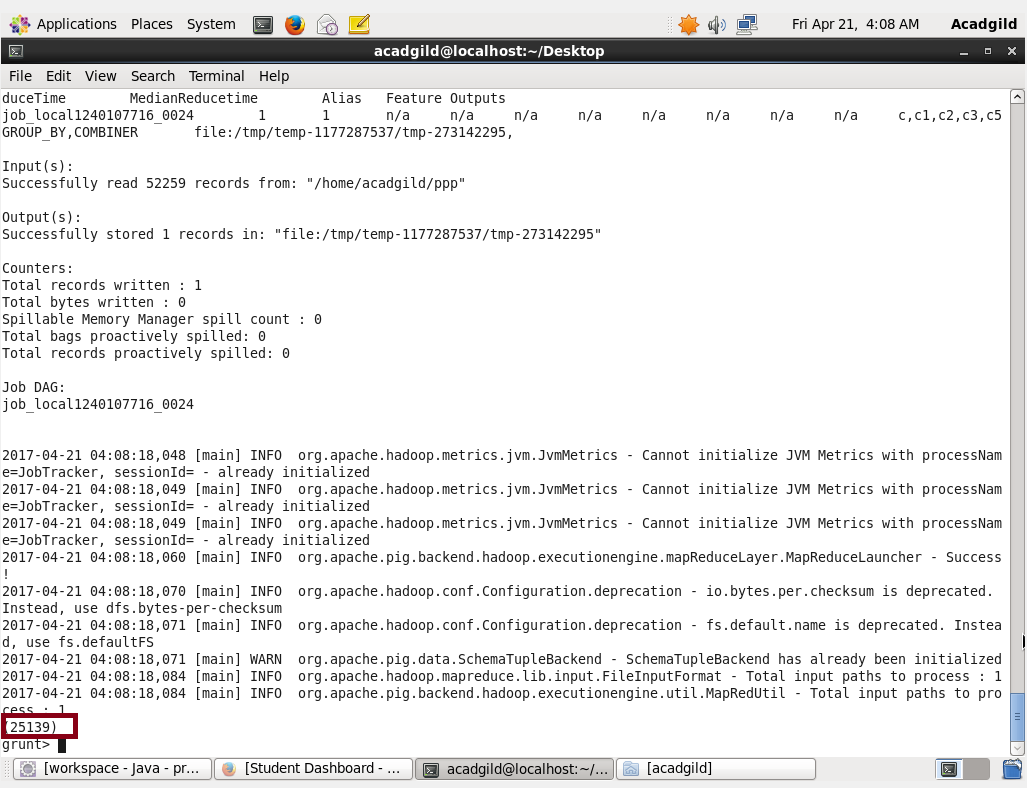
1. **WRITE A PIG SCRIPT TO FIND NO OF COMPLAINTS WHERE CONSUMER FORUM FORWARDED THE COMPLAINT SAME DAY THEY RECEIVED TO RESPECTIVE COMPANY**

LOGIC:

1. LOAD command - Will load the dataset as grunt shell..
2. Data sheet is filtered to check whether column 1 and 12 are same.
3. GENERATE command with 1 and 12 will generate only column 1 and 12.
4. GROUP ALL command – used to group all.
5. Generating the count using COUNT command

****

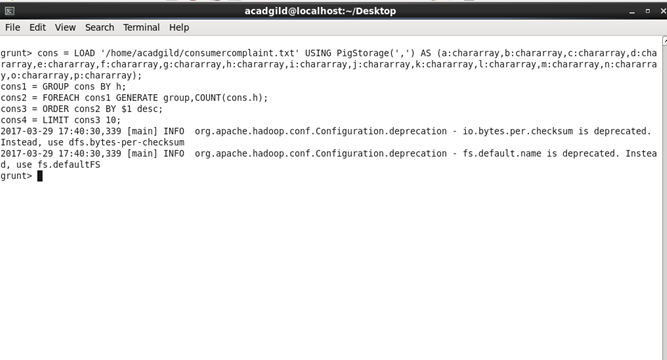
**OUTPUT:**

****

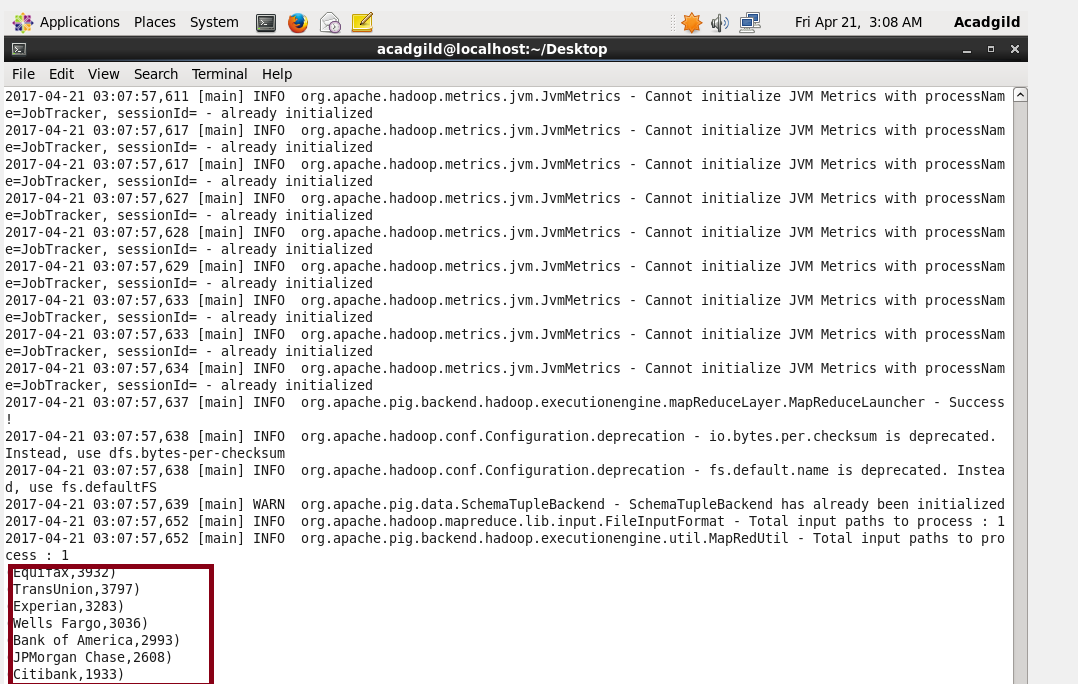
1. **WRITE A PIG SCRIPT TO FIND LIST OF COMPANIES TOPPING IN COMPLAINT CHART (COMPANIES WITH MAXIMUM NUMBER OF COMPLAINTS)**

**LOGIC:**

1. LOAD command - Will load the dataset as grunt shell.
2. Grouping the Dataset by Company name column.
3. COUNT command – will generate count .
4. Descending and ORDER will be used to order and count those in descending order using ORDER BY command.
5. LIMIT command – will display the top 10 companies.

**

**OUTPUT:**

****

1. **WRITE A PIG SCRIPT TO FIND NO OF COMPLAINTS FILED WITH PRODUCT TYPE HAS "DEBT COLLECTION" FOR THE YEAR 2015**

**LOGIC:**

1) Filter command is used to find whether any two columns contain the same debt collection.

2) Year (yyyy) is generated using GENERATE and SUBSTRING command

4) year =”2015” will filter the contents.

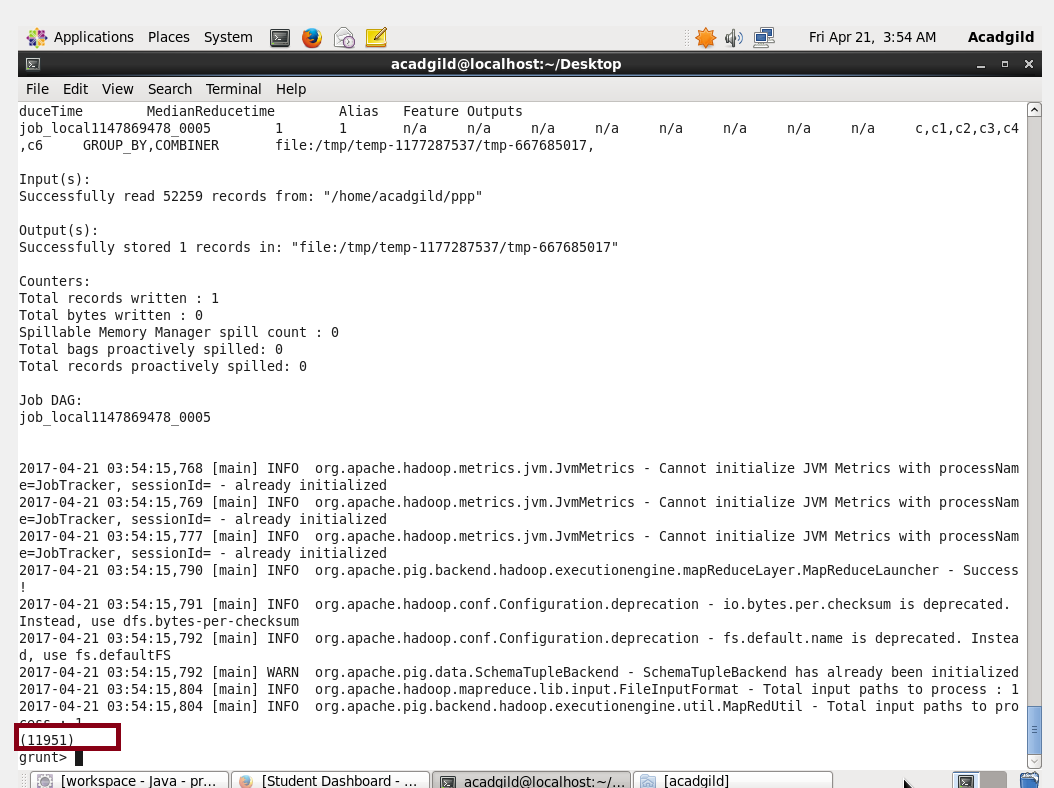
5) GROUP BY command –will group Debt collection.

6) COUNT - command will count the output .

7) Displaying only the count as output.

****

**OUTPUT:**

****