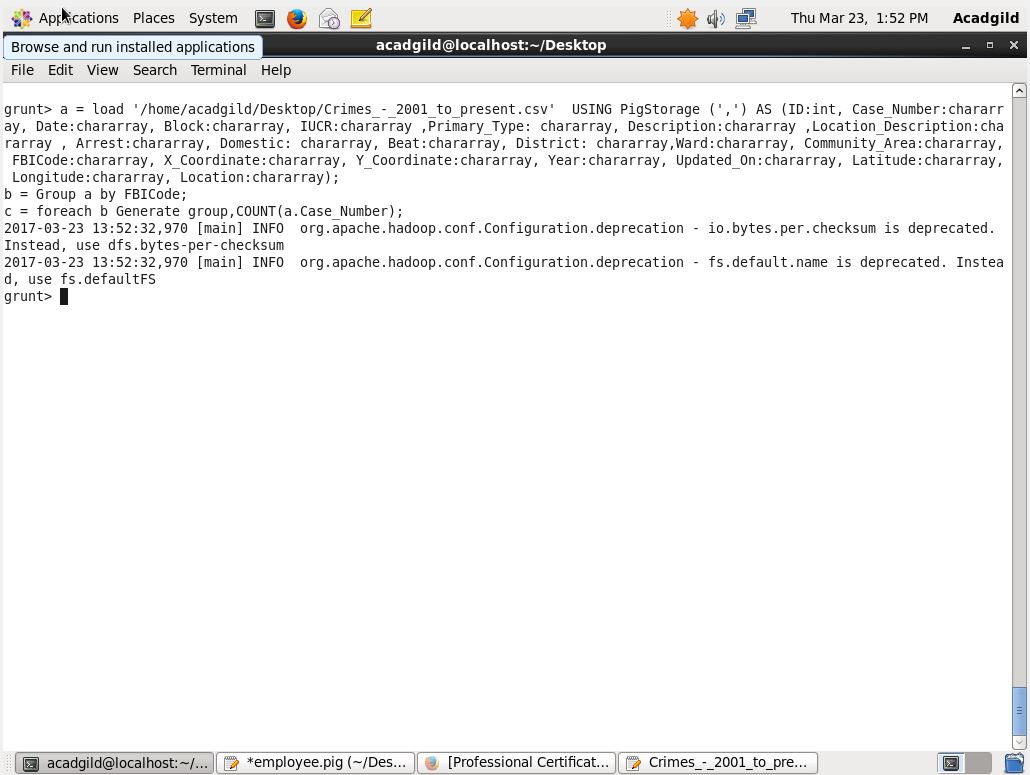
1. Write a mapreduce and pig program to calculate the number of cases investigated under

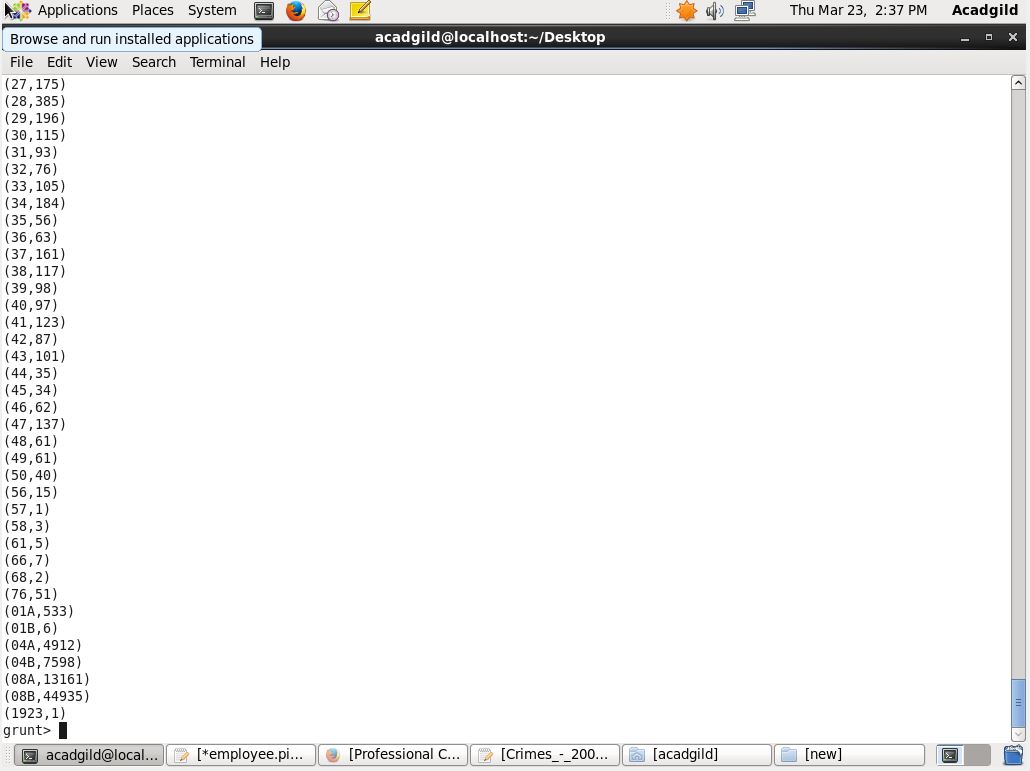
each FBI code

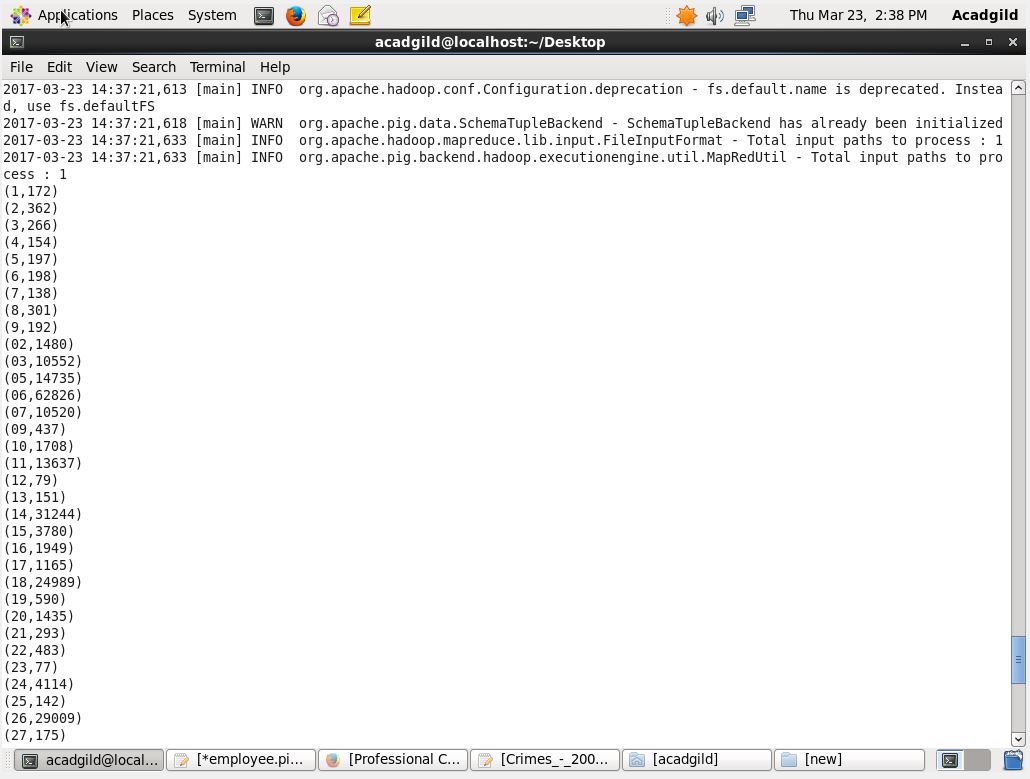
Here, the code is for the pig program. Using the load command we have loaded the dataset and it can be dumped or described using the command on ‘a’. now as we have to find the count of the number of cases under each FBI code we have grouped using FBI code. Then using the count we have generated number of cases investigated under each FBI code using COUNT command.

By dumping ‘c’ we can have the FBI code and the number of cases investigated under each code.



As we have dumped the ‘c’ The answer is as follows: FBI code, number of cases investigated.

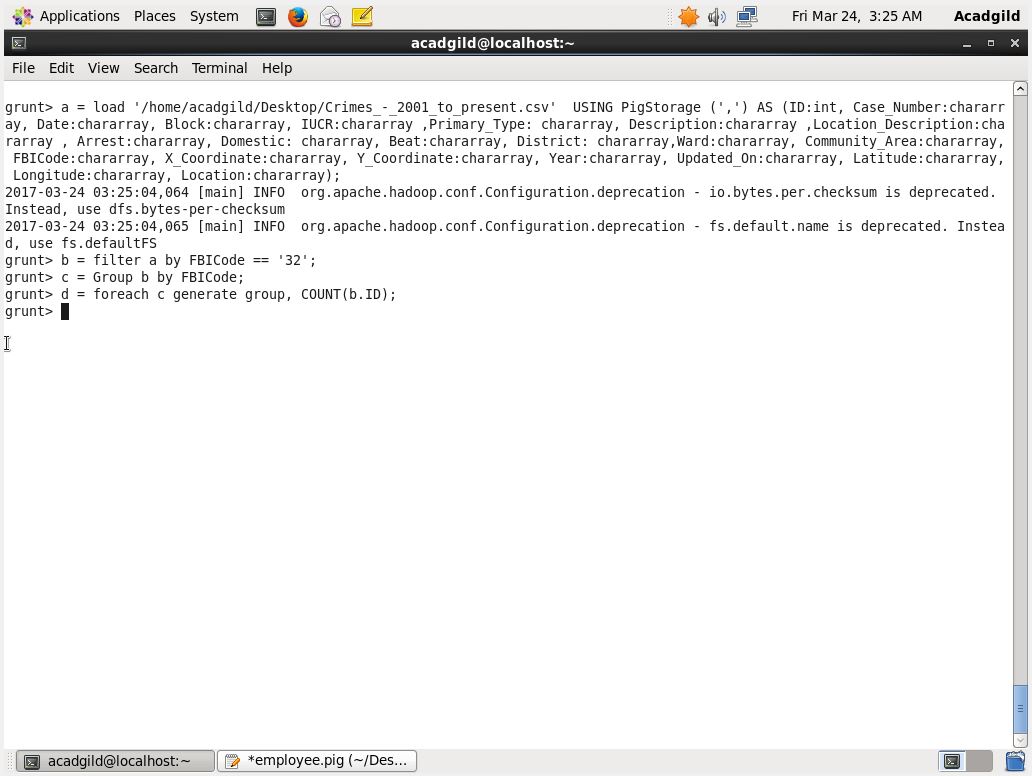




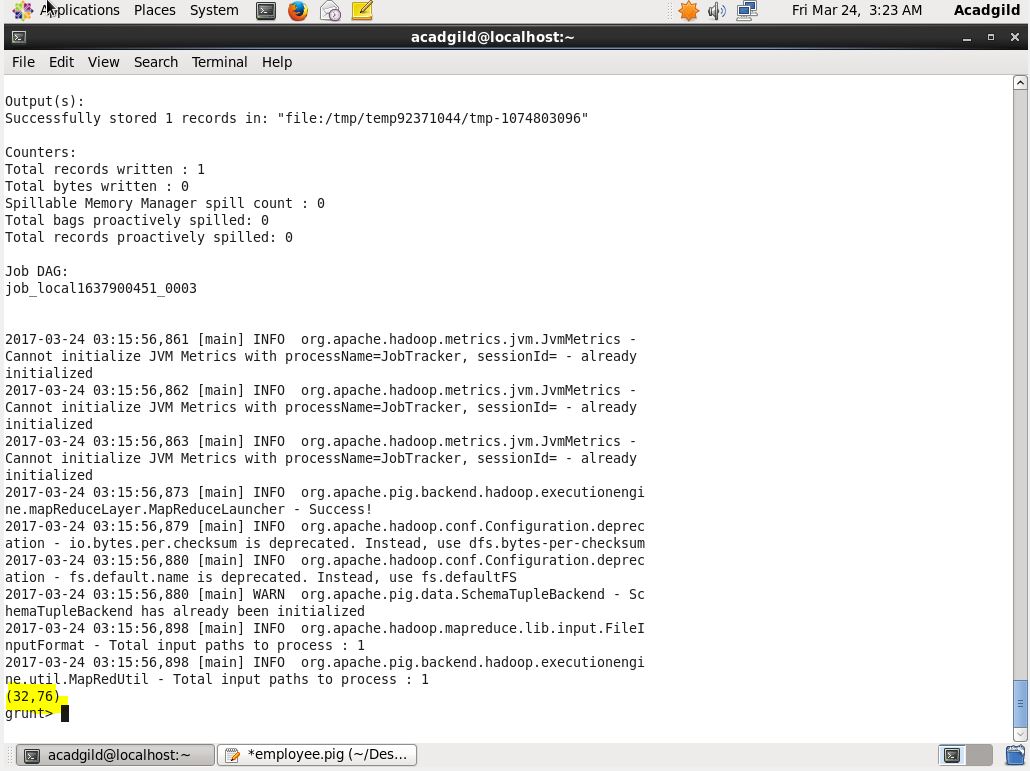
2. Write a mapreduce and pig program to calculate the number of cases investigated under FBI

code 32.

The pig script is written in pig and here we are supposed to find the number of cases that have been investigated under FBI code 32. For that we have first loaded the dataset using the load command. Then the filter operation is performed to fetch the data of FBI code 32. Grouping is done after the filter operation is done. Then after grouping is done we need to find the count of the cases investigated under the FBI code 32.



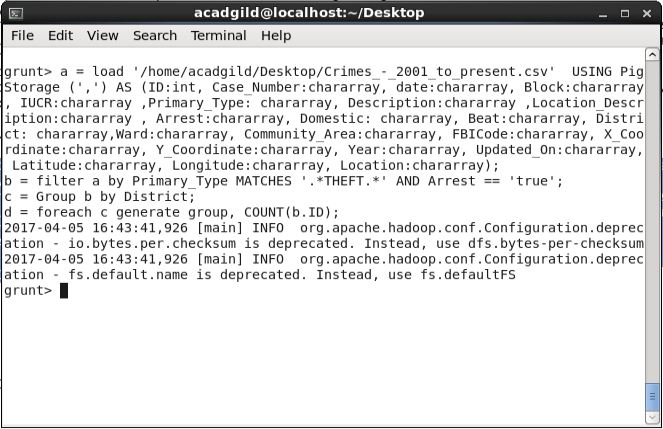
The number of cases that are investigated under the FBI code 32 are 76. This can be obtained by dumping the variable ‘d’ where the count and the code is stored.



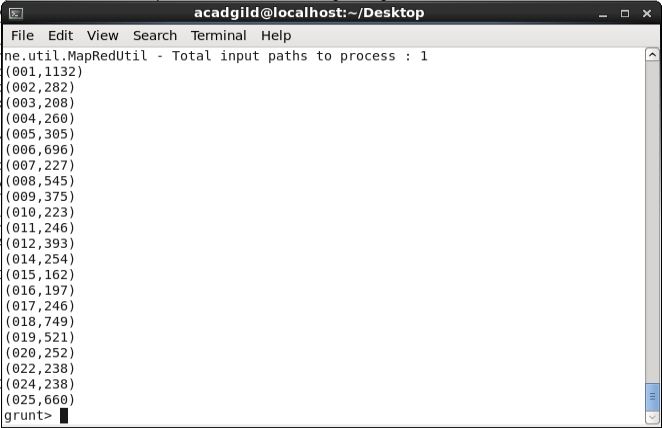
3. Write a mapreduce and pig program to calculate the number of arrests in theft district wise.

The pig script is used of the dataset to find the number of arrests under the theft district wise.

Here as we load the dataset the filter operation on that dataset is done as to get the data which are of theft and the arrest is done i.e. true. After the filter we group it using the district id as we need to find the count district wise. Once the grouping is done the count is generated for each district and is stored in variable ‘d’. To fetch the result we need to dump the variable ‘d’.



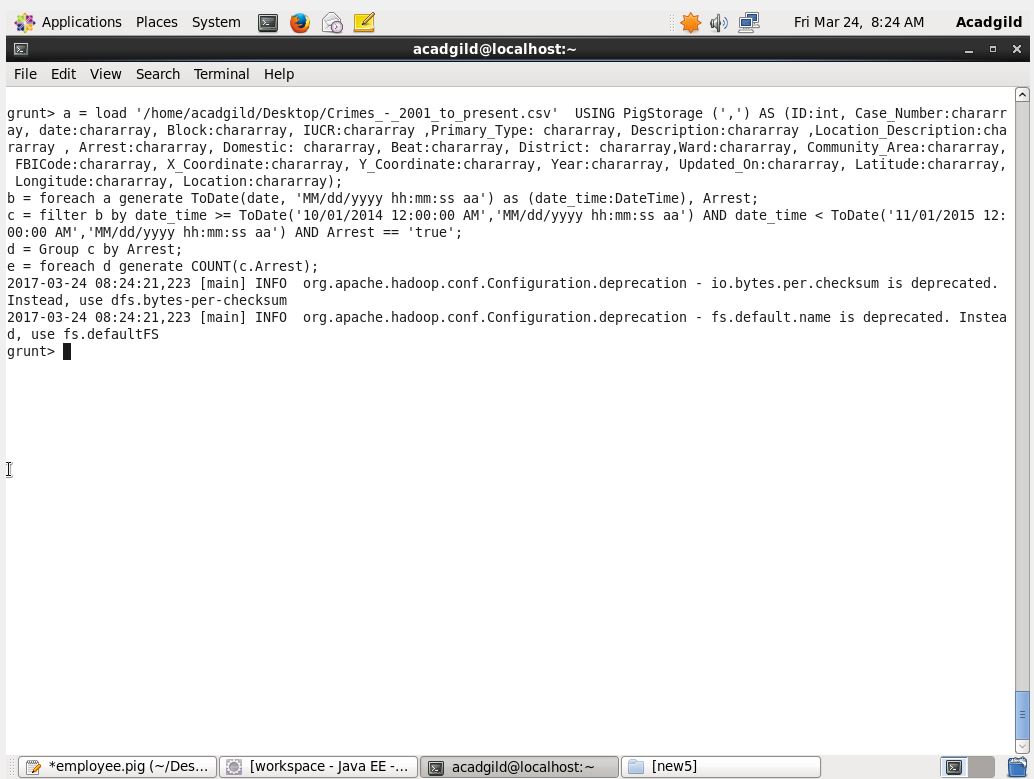
As we dump the result is shown as district id , number of arrests that are done.



4. Write a mapreduce and pig program to calculate the number of arrests done between

October 2014 and October 2015.

The pig script written here is used to find the number of arrests that are being done between October 2014 and October 2015. We have included the starting and the end seconds also of the days i.e. 1st of October, 2014 12:00:00AM to 31st of October,2015 11:59:59 PM. To get required result we first loaded the data and then the date column is converted date time format using todate command. Then the filter operation is performed to get the data between the required dates from dataset as well as the condition for the arrest is also checked if it is true. Grouping is done on the basis of the arrest. The count is then generated and is stored. To get the count we dumped the variable ‘e’.



The answer for the question is 63173 the total arrest that are conducted between the given duration:

