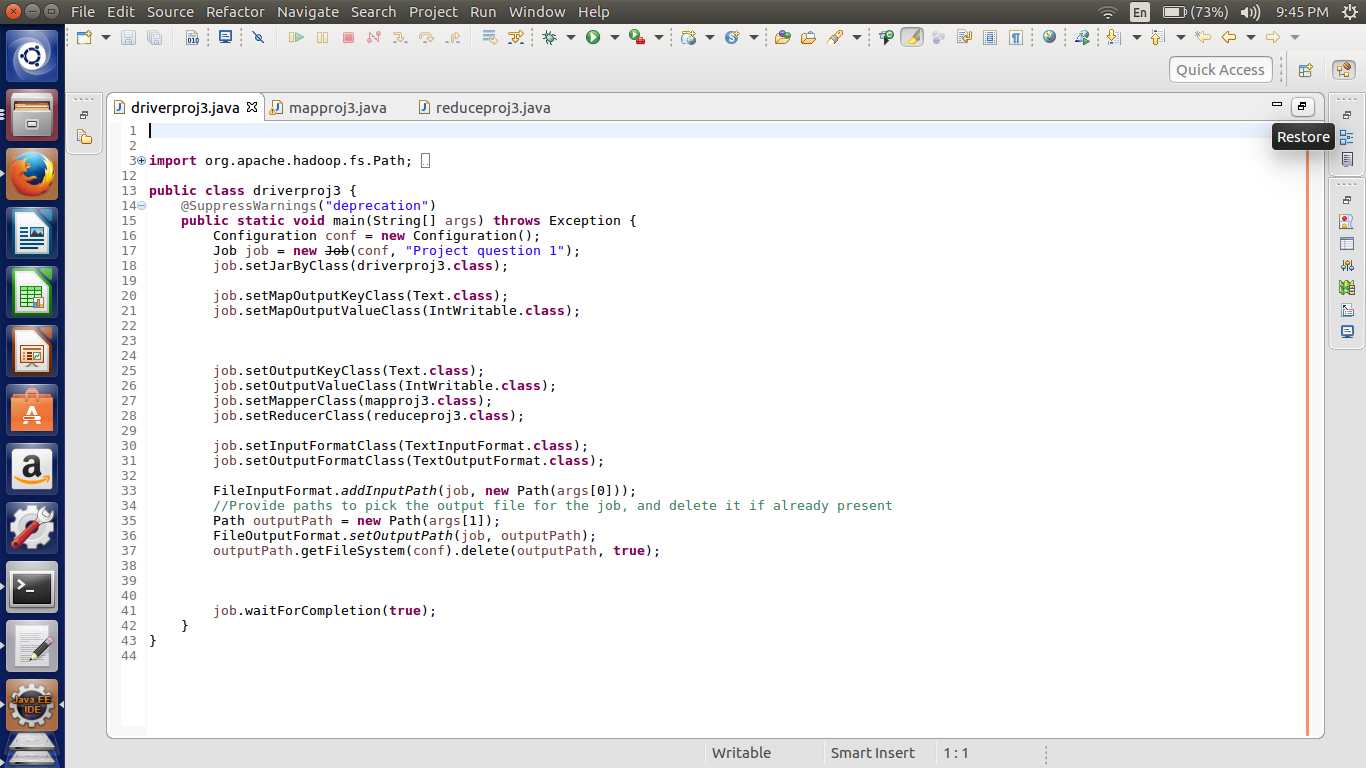
3. Write a mapreduce and pig program to calculate the number of arrests in theft district wise.

//Driver class

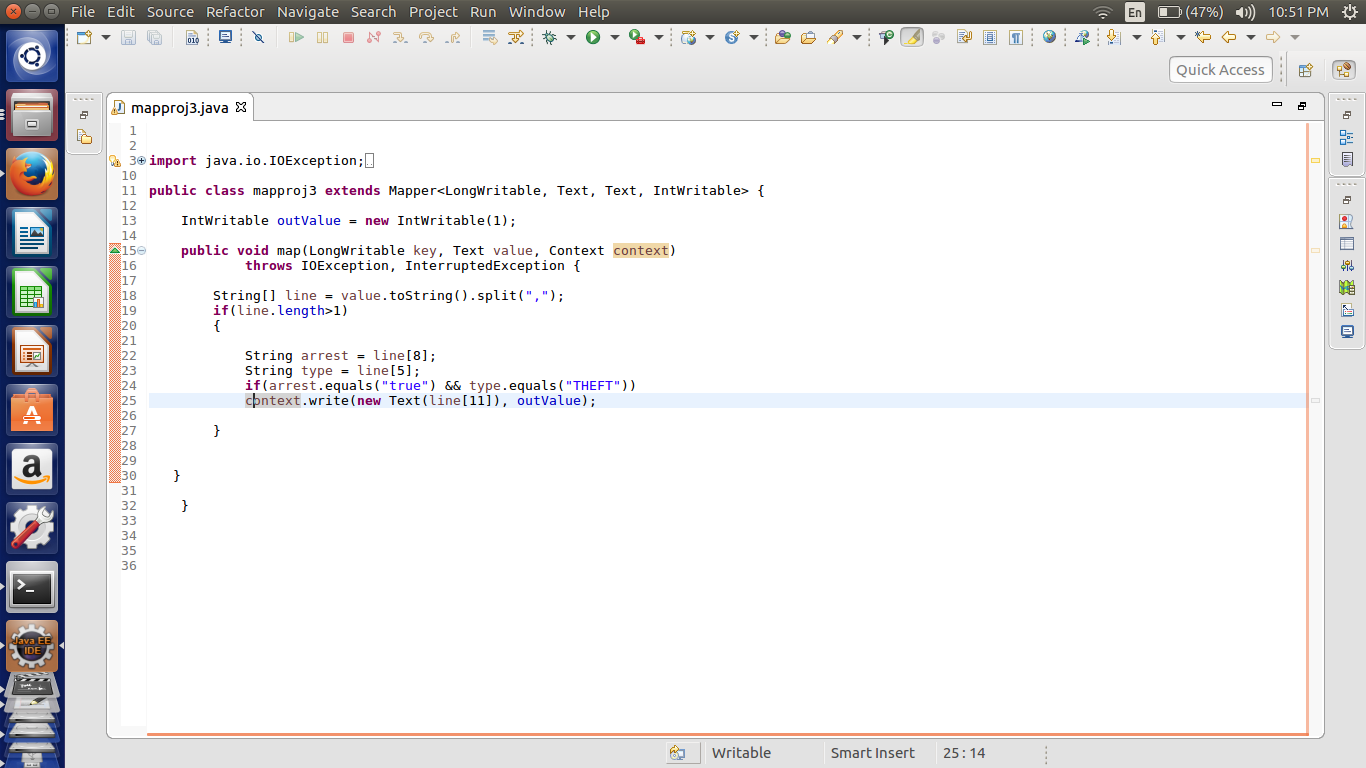


Here we are setting the job configurations and in the first question we can achieve the results by doing the map and reduce job hence setting the in the job configurations about the map class as well as the reducer class.

Here I have taken the input and output format as the “Text Input Format class”.

Third last line of the code is written for the deleting the output file if it already exists in the HDFS.

//Mapper

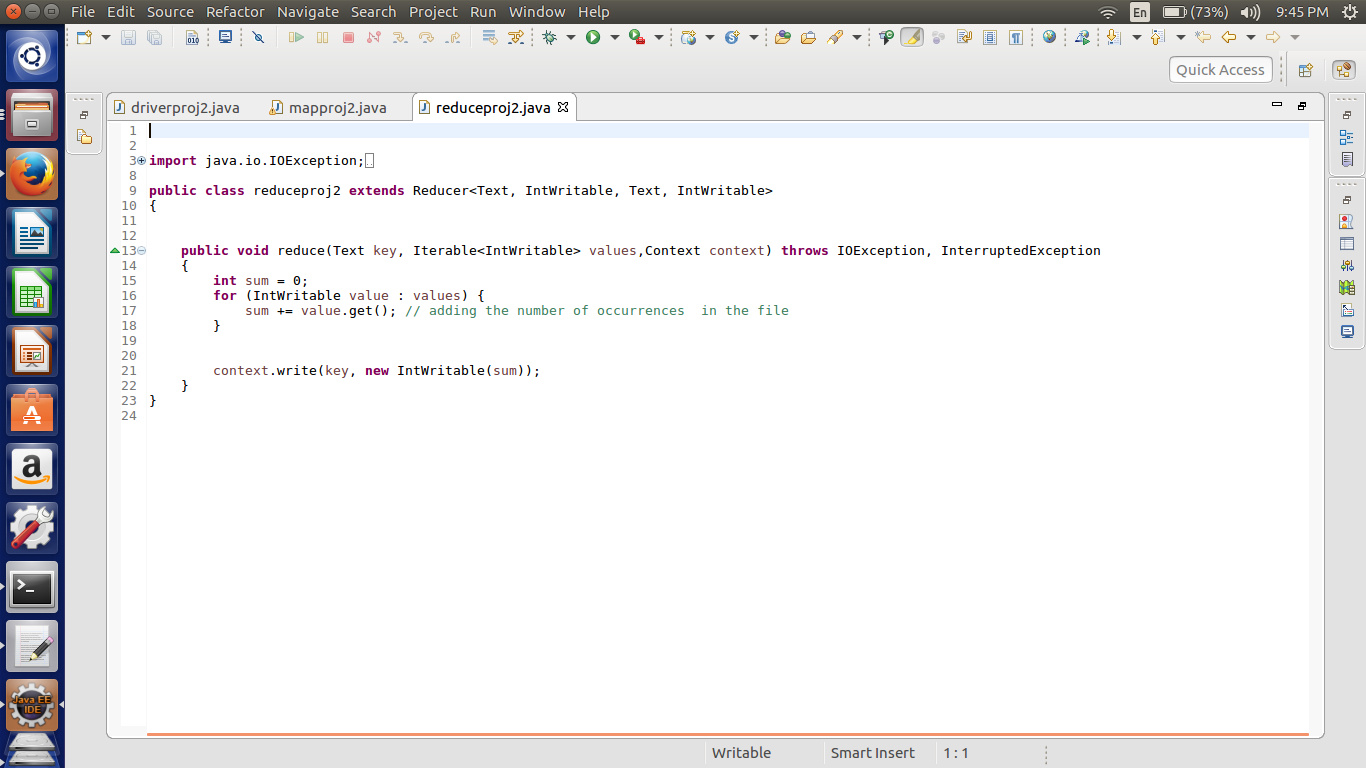


Here we are checking the condition that the arrests is true and the type includes THEFT only. Other forms of the **THEFT are not considered.**  If the other forms of the theft are also to be considered the one should use the String MATCHES and implement the same code.

If the condition is true then we are passing the district and the intwritable values as 1.

This Intwritable values will be added in the reducer to get the total arrests done under the case of the theft, District Wise.

//Reducer

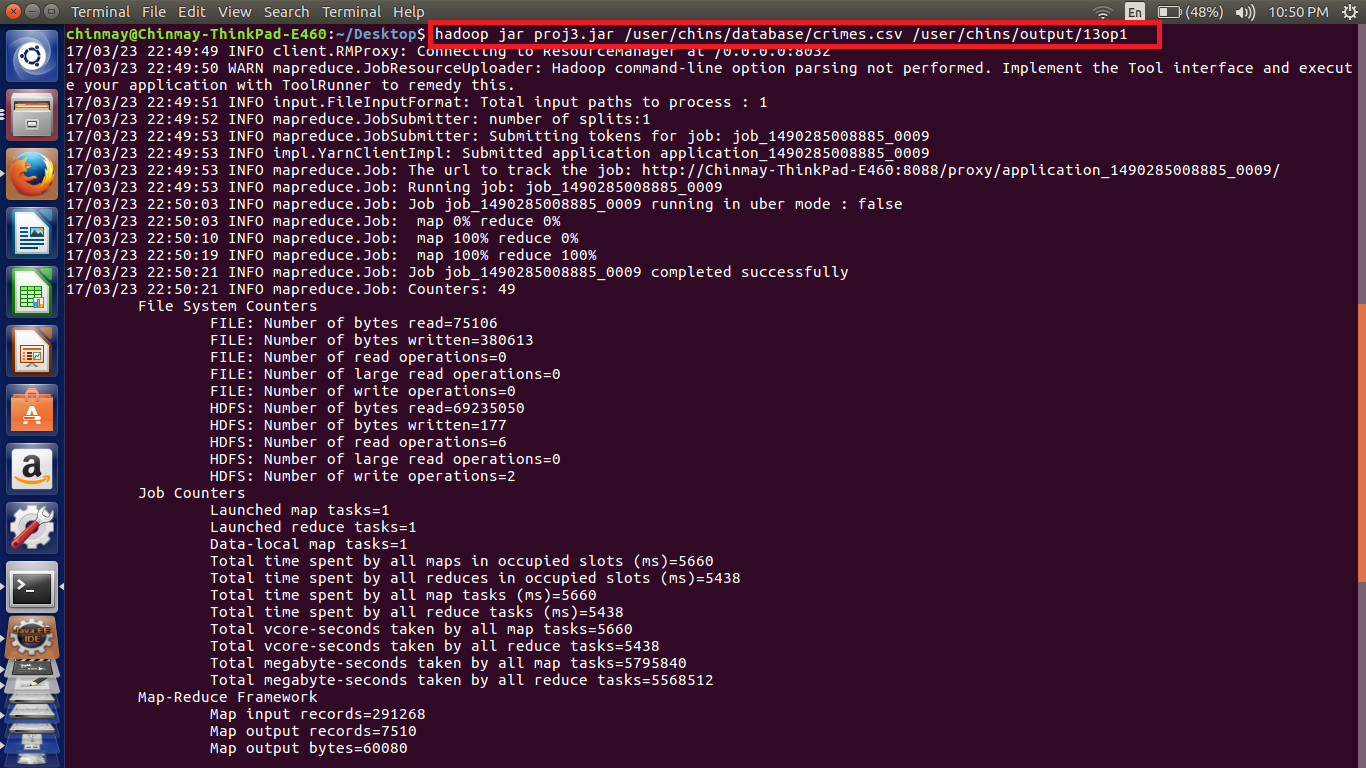


Reducer is simple implementation of the adding up all the values related to each key. Here the values only one.

As for each occurrence of the condition the value one is passed so that we total in each district.

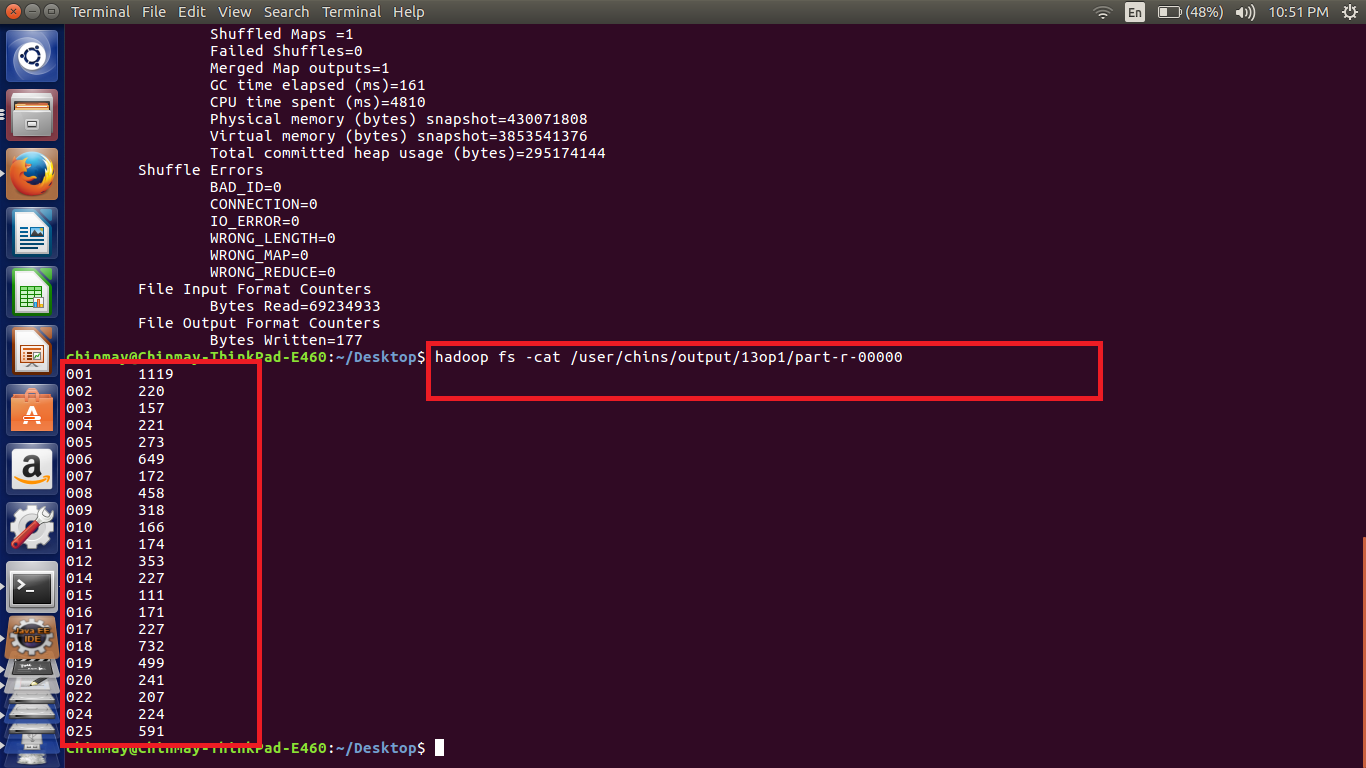
We have passed the District as the key.

// Implementation



Here we can see the implementation the code using the highlighted command in which the input and output file paths are mentioned.

//Output Mapreduce



PIG Implementation

a = load '/home/chinmay/Desktop/crimes.csv' USING PigStorage(',') as (id:int,caseno:chararray,date:chararray,block:chararray,IUCR:int,type:chararray,desc:chararray,locationdesc:chararray,arrest:chararray,dome:chararray,beat:int,dist:int,ward:int,Coarea:int,code:chararray,xco:int,yco:int,year:int,Upd:chararray,latitude:int,longitude:int,location:chararray);

b = FILTER a by arrest=='true'AND type=='THEFT';

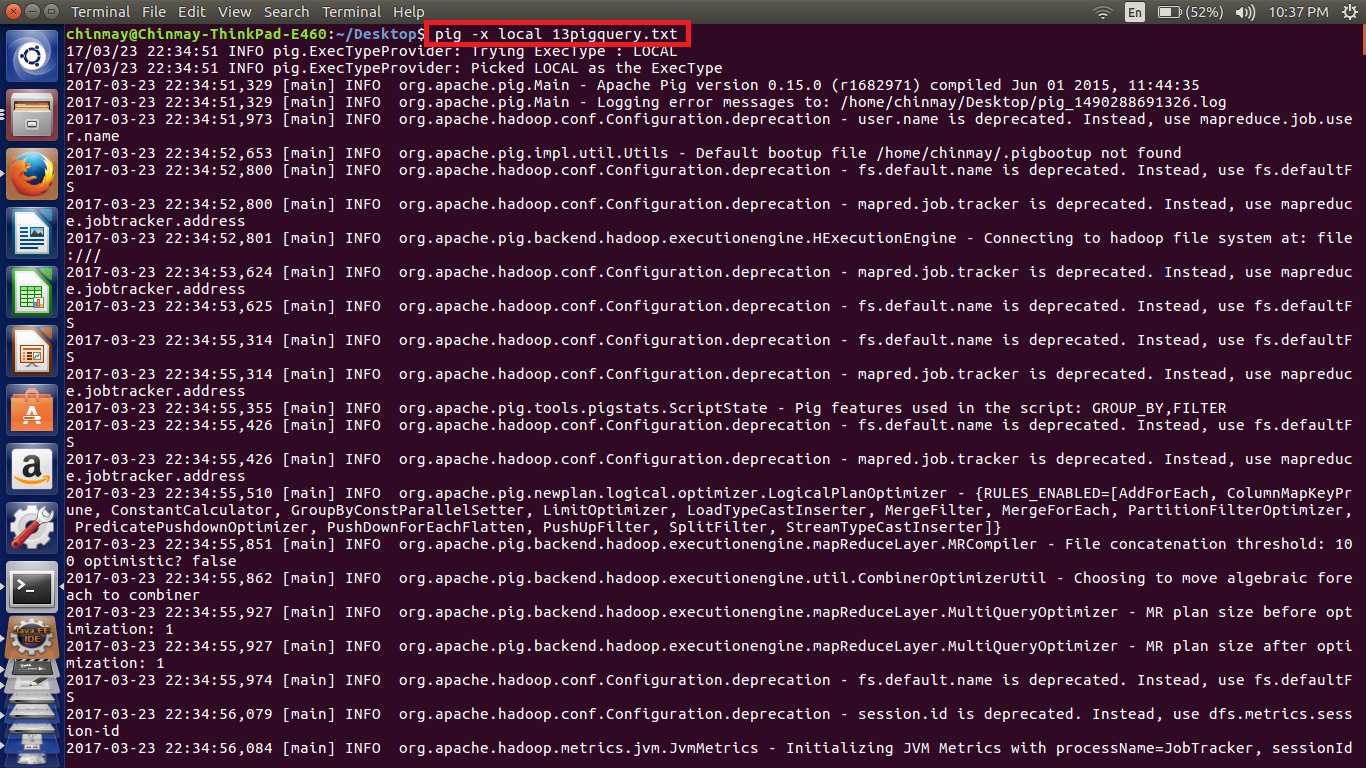
c = GROUP b by dist;

d = FOREACH c GENERATE group,COUNT(b.arrest);

dump d;

Note – File used for the execution is also attached.

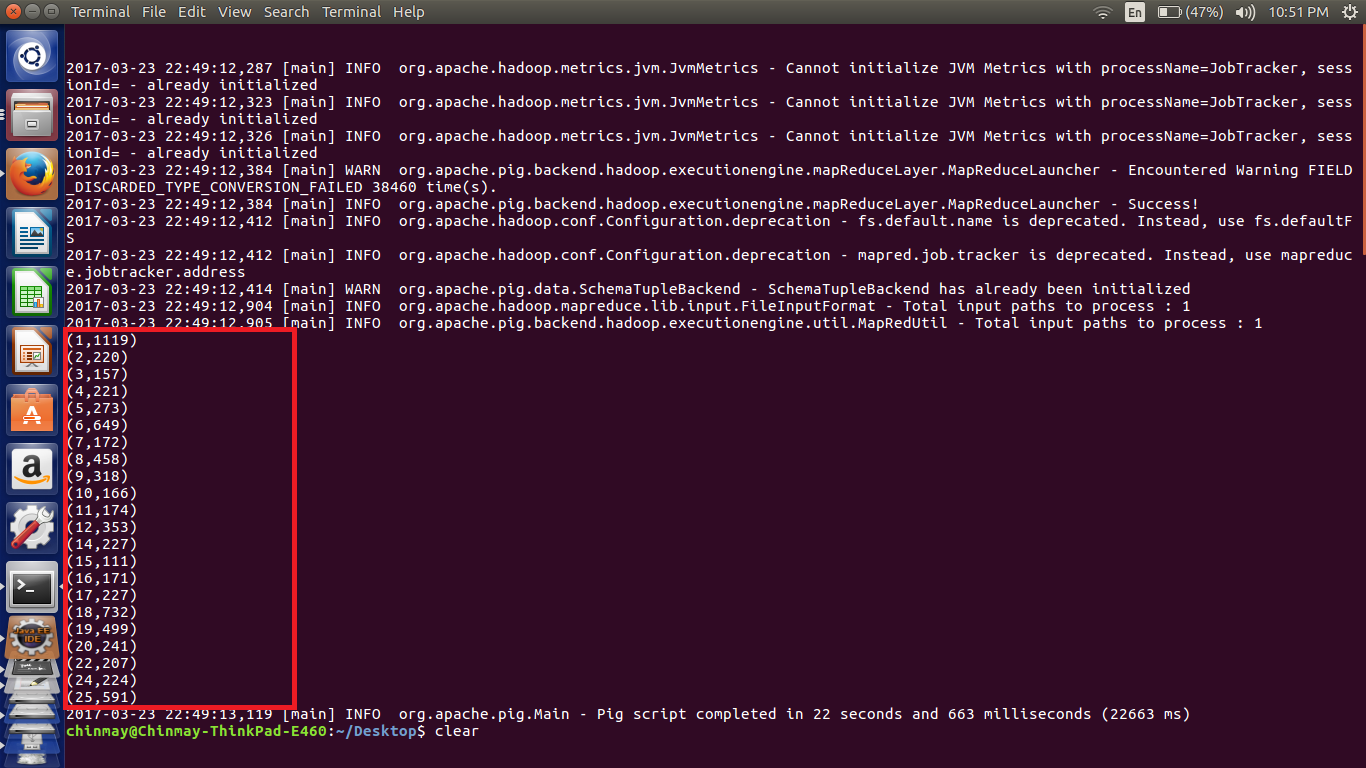
//Implementation



Here in the pig implementation we are loading the data according to the respective datatypes and then filtering using the AND operator to check two conditions of Arrests and the theft only

Then we have grouped by District so as to get the results according to the district. And then counting the number of arrests will give us the no of arrests with that condition.

//Output



We can see the output by the simple mapreduce program as well as the pig implementation are the same.