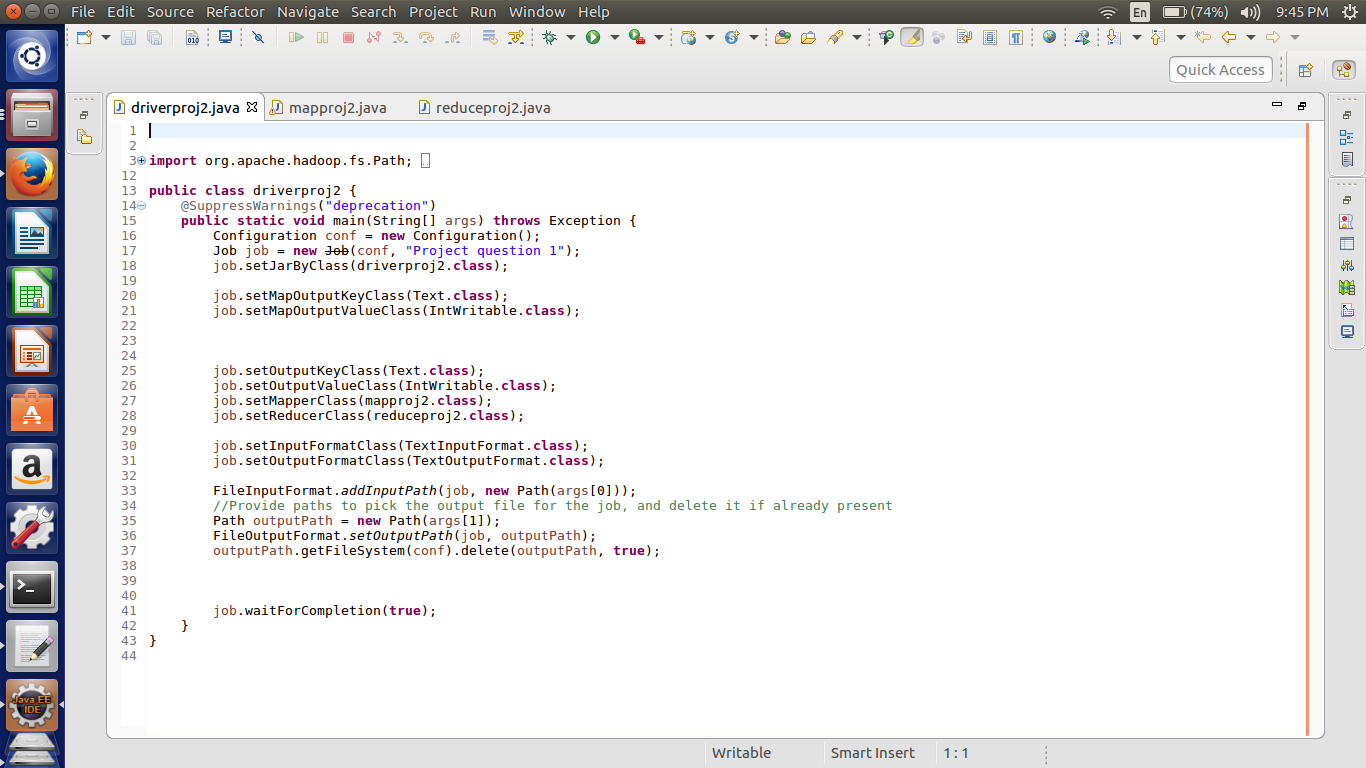
2. Write a mapreduce and pig program to calculate the number of cases investigated under FBI code 32.

//Driver class

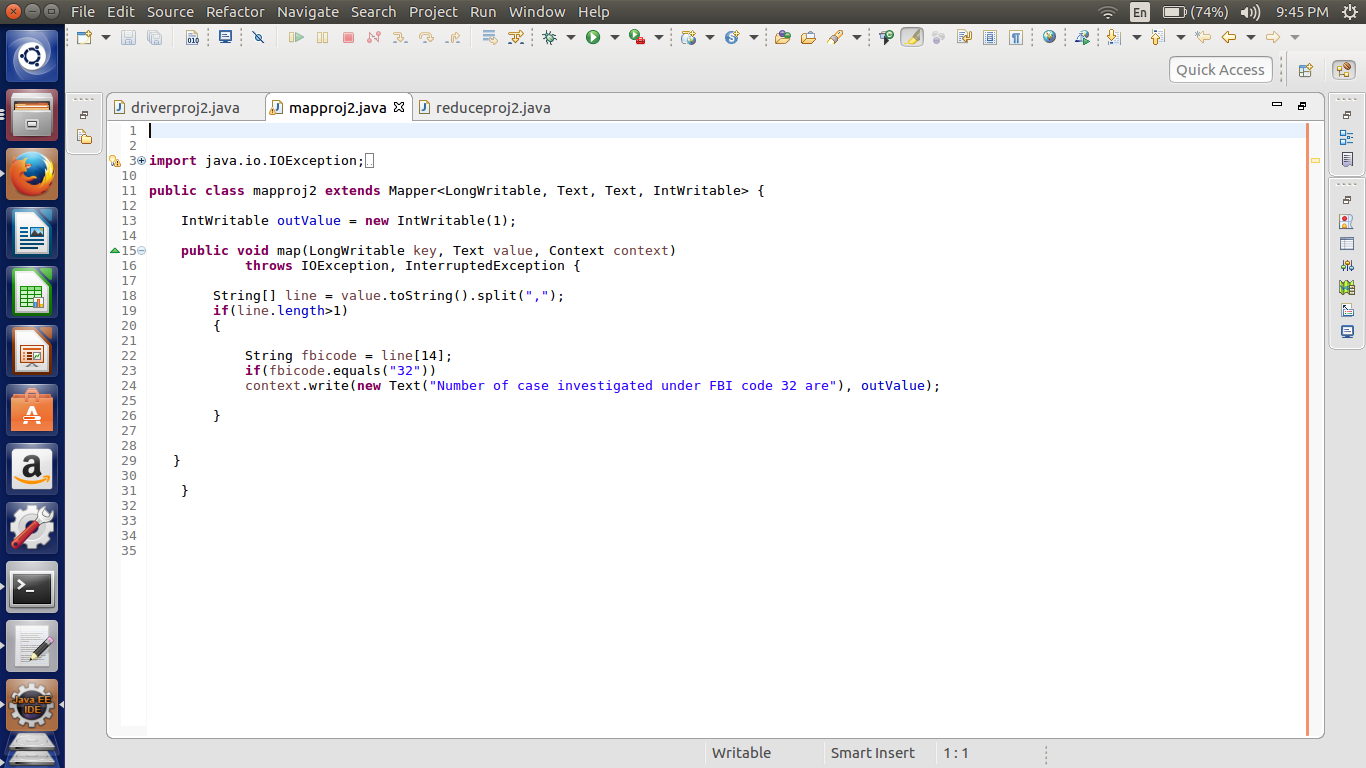


Here we are setting the job configurations and in the first question we can achieve the results by doing the map and reduce job hence setting the in the job configurations about the map class as well as the reducer class.

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

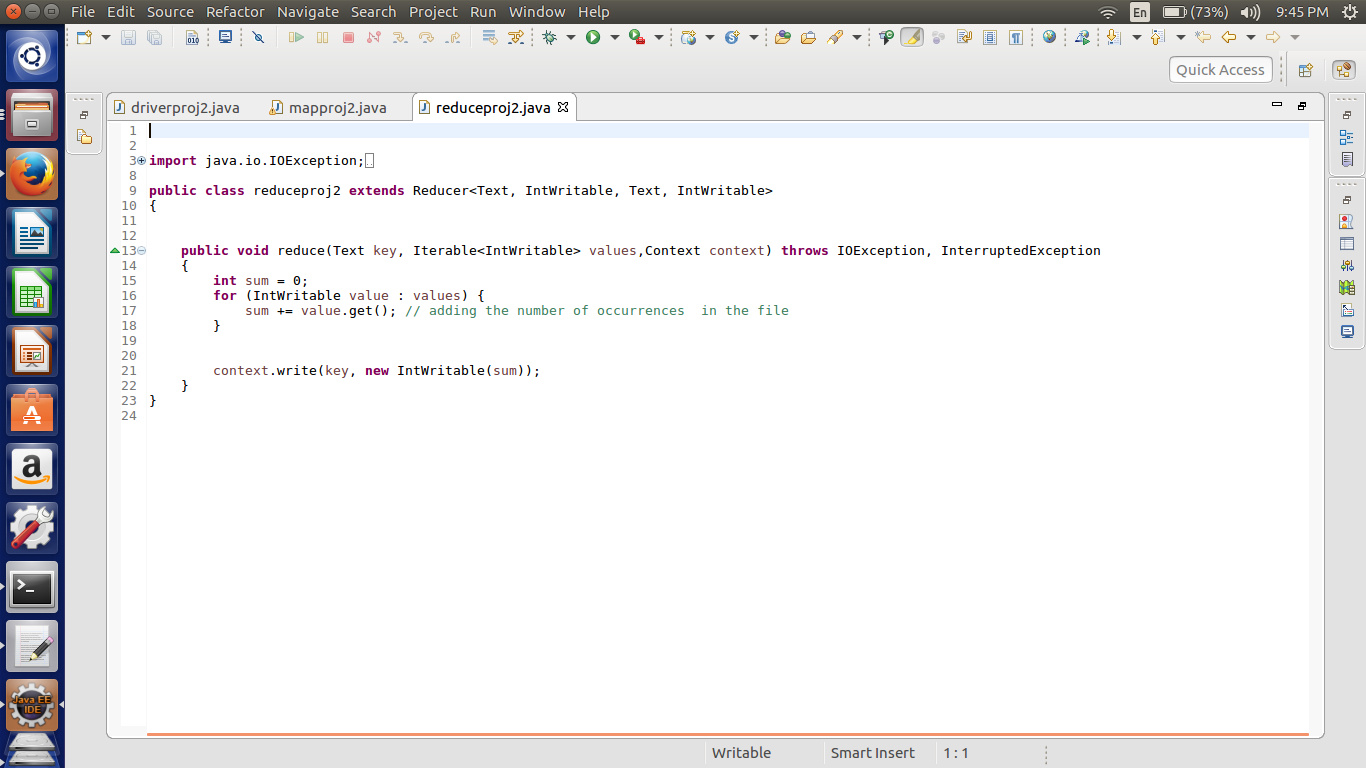
Third last line of the code is written for the deleting the output file if it already exists in the HDFS.

//Mapper Class



Here passing the key as the constant text as “Number of case investigated under the FBI code 32 are” so the key will remain constant. Now the value of the key is in intwritable format and is fixed as one so whenever there is an occasion of the code 32 the output value one will be registered along with the constant key.

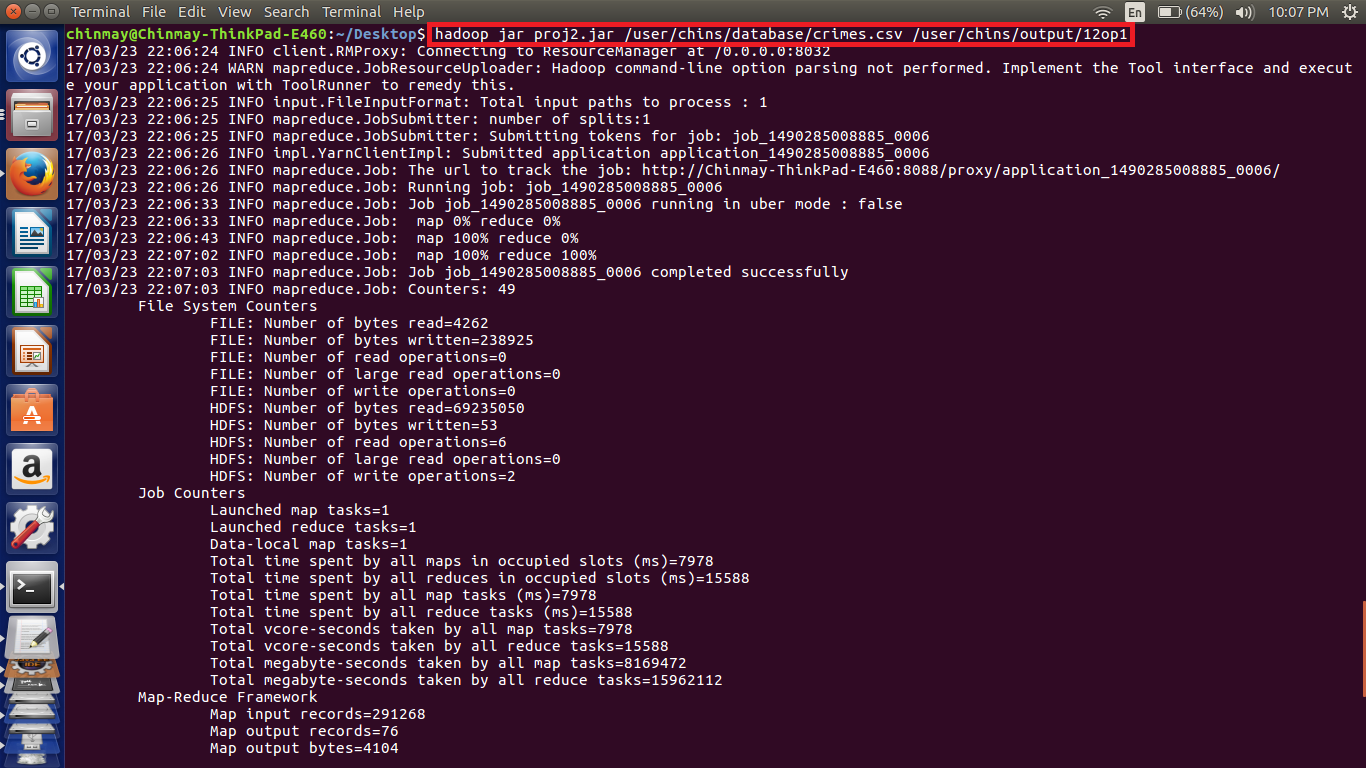
//Reducer Class



In the reducer the summing up the values which are related each key and passing the key and the sum as the value in the intwritable format.

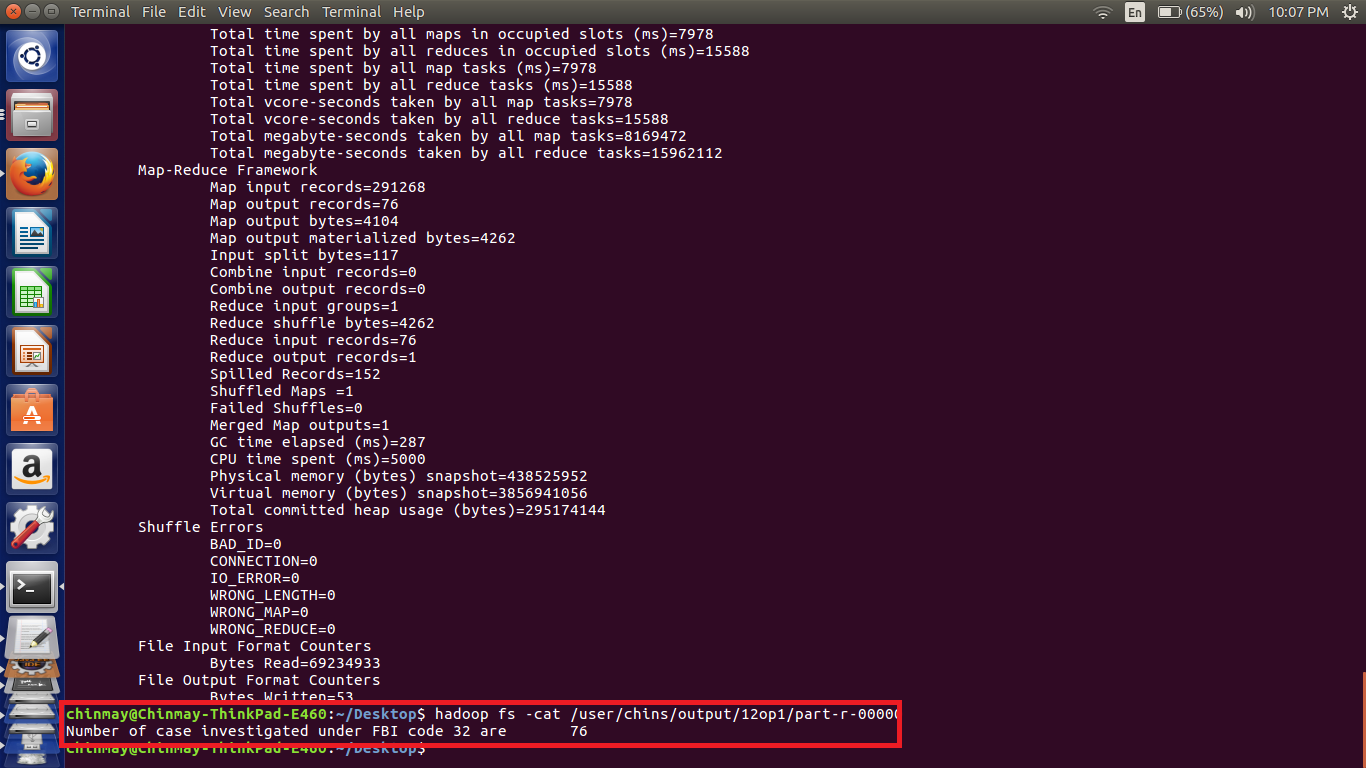
Here as we have the single key all the values related to that key will be added.

//Implementation



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

//Output



Here we can see the output as key and value and that the 76 cases are under the FBI code 32.

PIG Implementation

//Pig Script

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 code=='32';

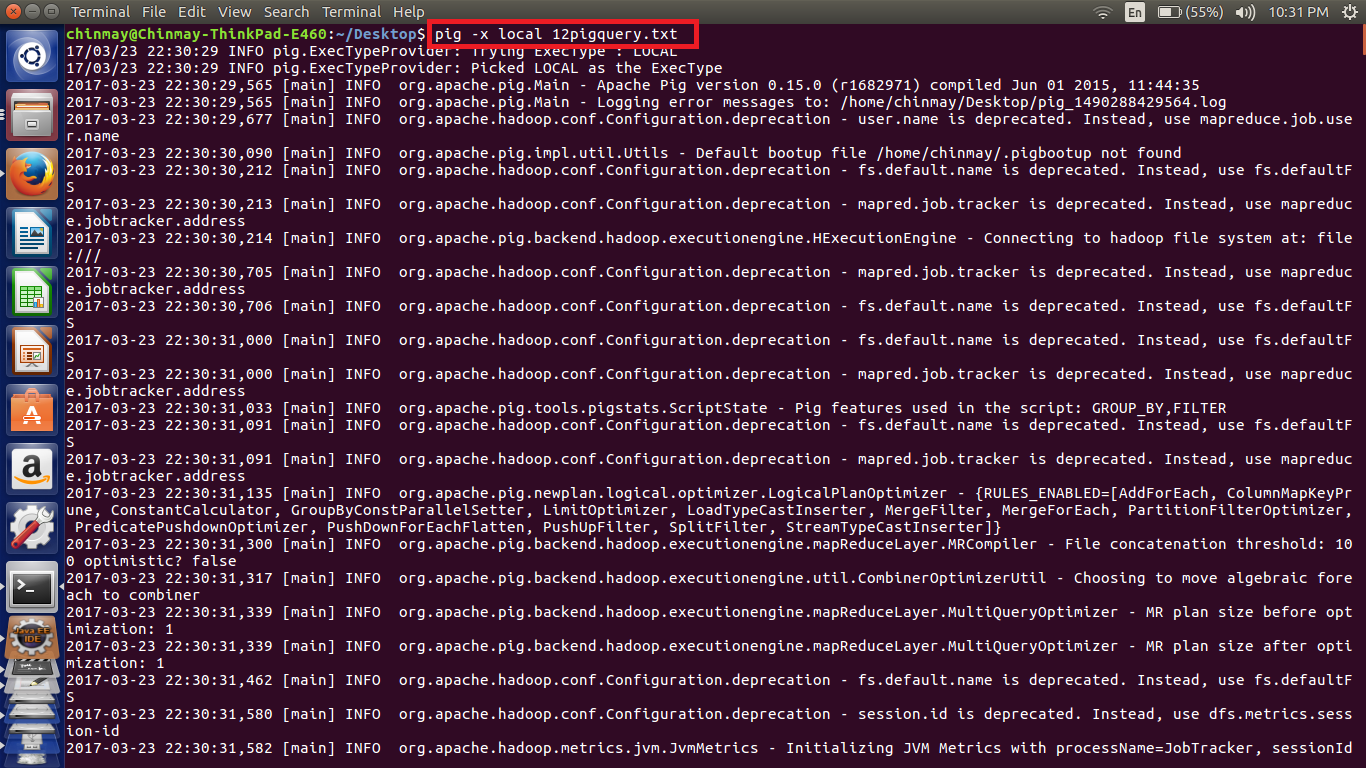
c = GROUP b by code;

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

dump d;

Note – File used for pig execution is also attached

//Implemetation

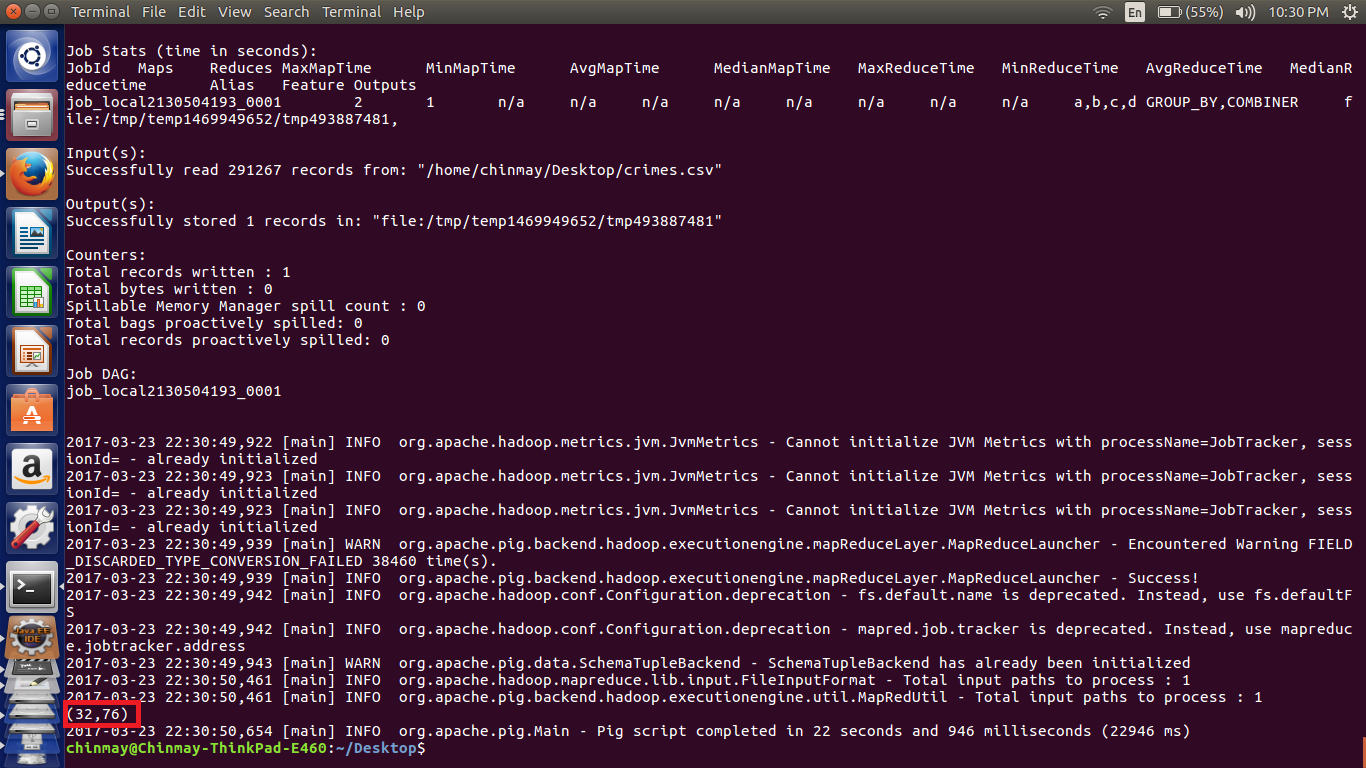


Here in the pig Script we have loaded the data according to the respective datatypes. Then filtered if the FBI code is 32.

Then grouped by the code. Here there is only one code.

Then counted the codes which will give us the number of cases.

//Pig Output



Here we can see the cases under the FBICode 32 is – 76 the output of the simple mapreduce program is same as that of the pig implementation.