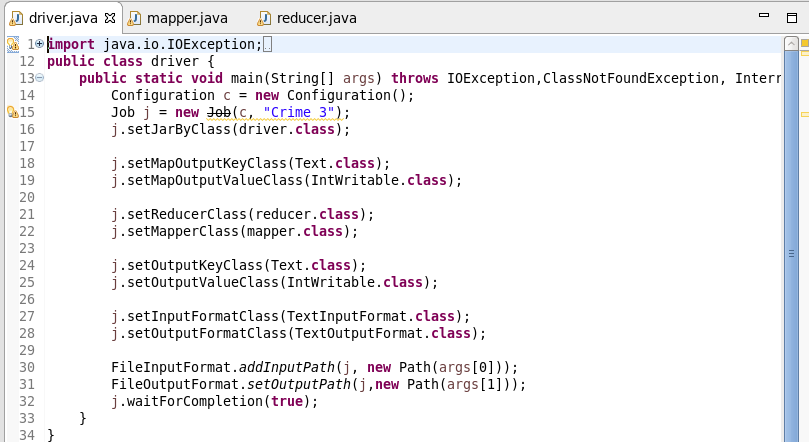
3. Write a Mapreduce and pig program to calculate the number of arrests in theft district wise

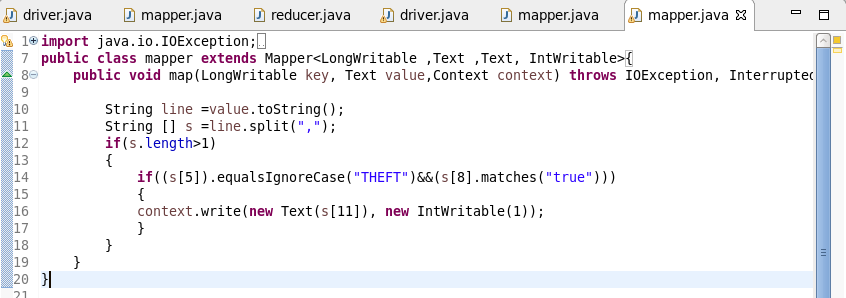
Driver Class:



Here we are setting the job configurations in the Driver Class and setting the Output and Input File Formats for the Mapper in the line 18 and 19. In the line 21 and 22 we are setting the Mapper and reducer class.

Here I have taken the input and output format as the “Text Format class” in the line 27, 28.

Mapper Class:

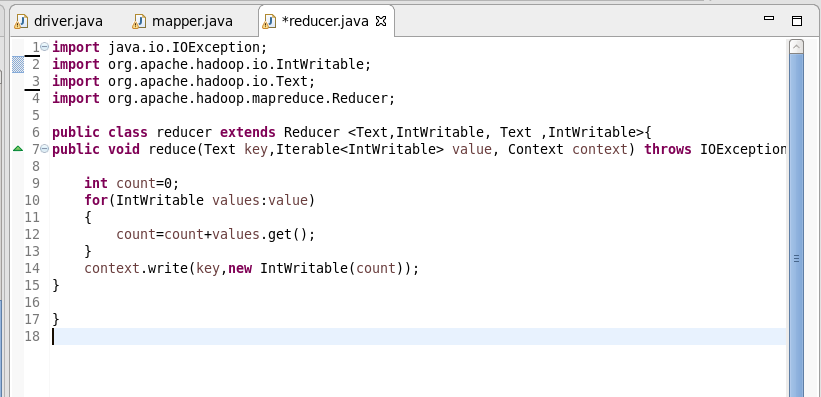


Here we are checking the condition that the arrests is true and the type includes THEFT only. Other forms of the **THEFT are not considered.**  If the other forms of the theft are also to be considered the one should use the String MATCHES and implement the same code.

If this 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.

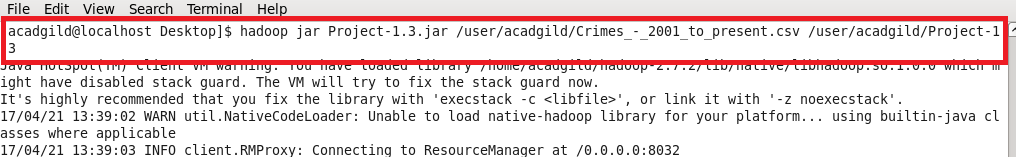
Reducer Class:



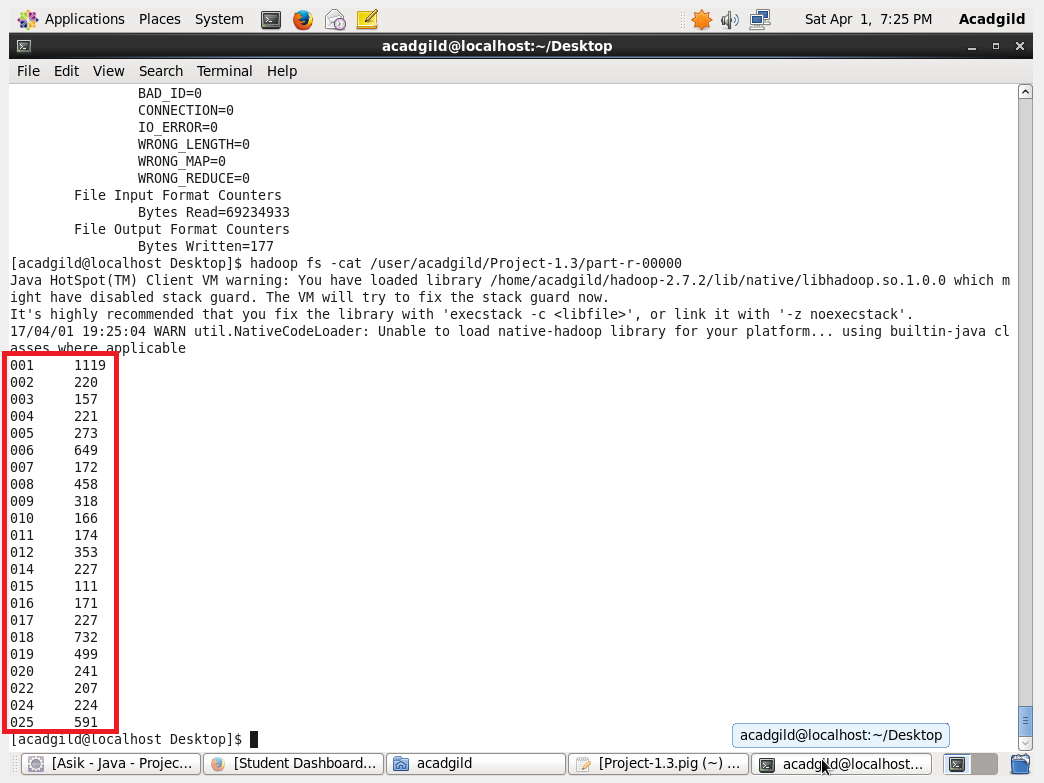
In the reducer we are summing up the values which are related each key and passing the key and the sum as the value in the intwritable format.

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

Running the Jar File:

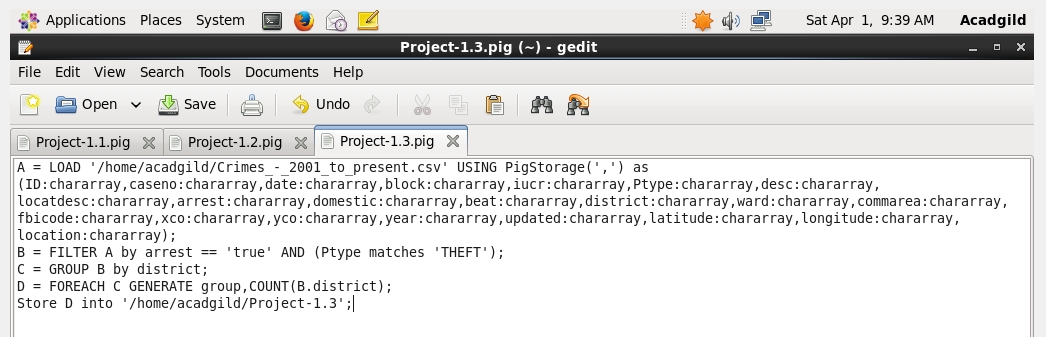


Output:



Pig Program:

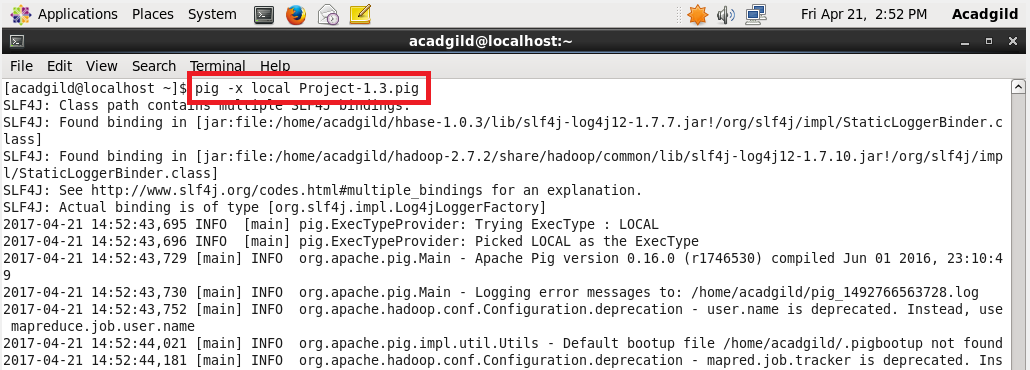
Pig Script:



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

Running the Script:



Output:

