PROJECT-1

**USA Crime Analysis**

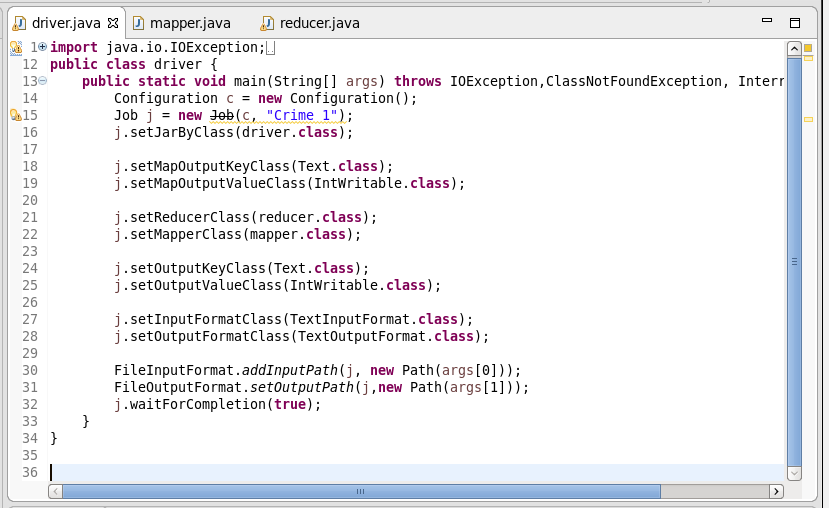
Introduction:

This dataset contains attributes related to crimes taking place in various areas like type of crime, FBI code related to that criminal case, arrest frequency, location of crime etc.

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

MapReduce Program:

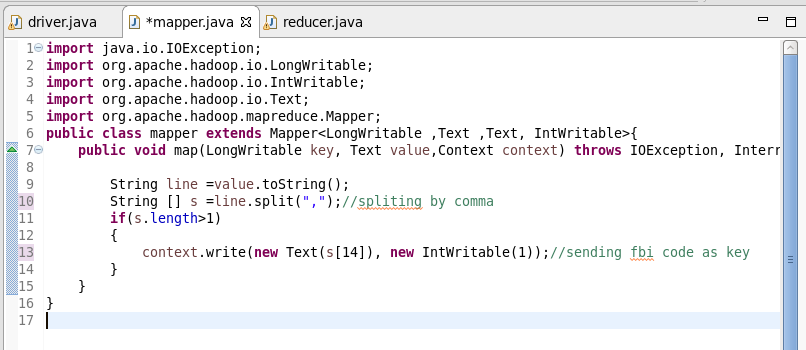
Driver Class:



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.

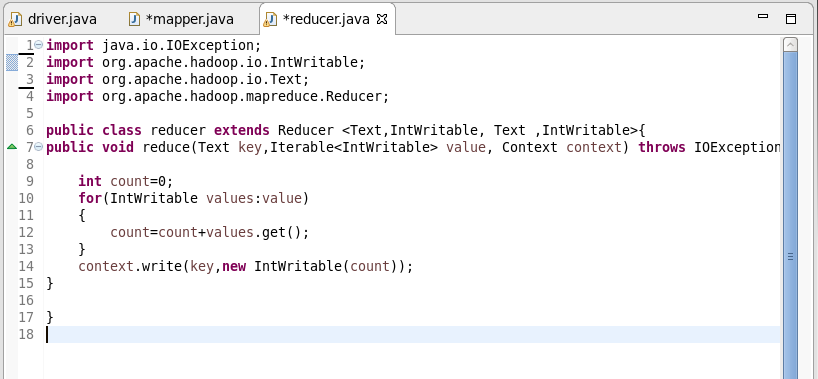
Mapper Class:



In the mapper class first converted the input line as the string then split using the delimiter as “ , ”.

Now consider the FBI code which is the 14th split sent it as the key and value as the Intwritable format which is one.

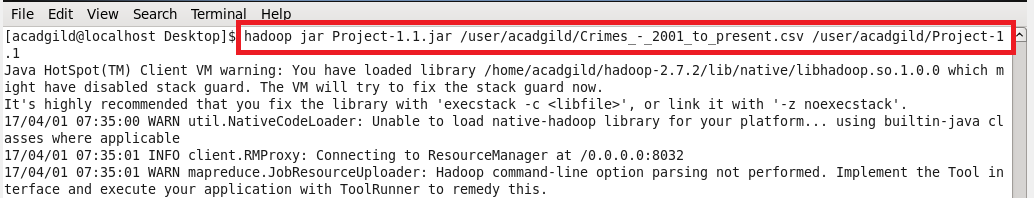
Reducer Class:



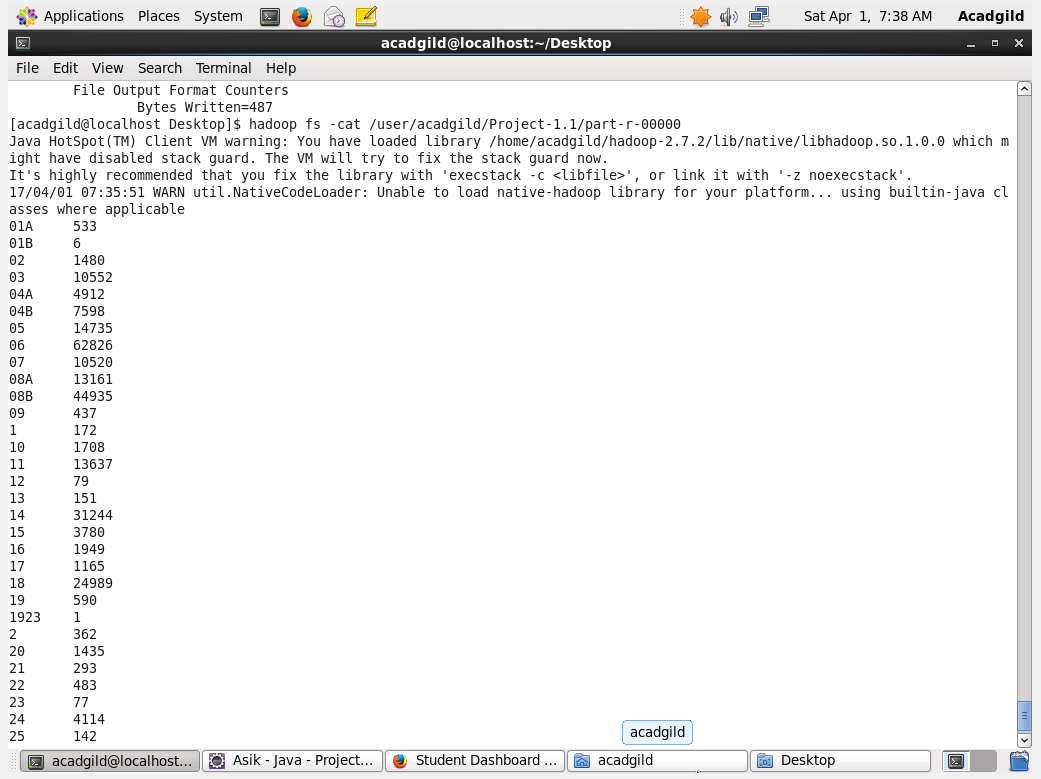
In the reducer we are summing up the values which are related each key and passing the key and the sum as the value in the intwritable format.

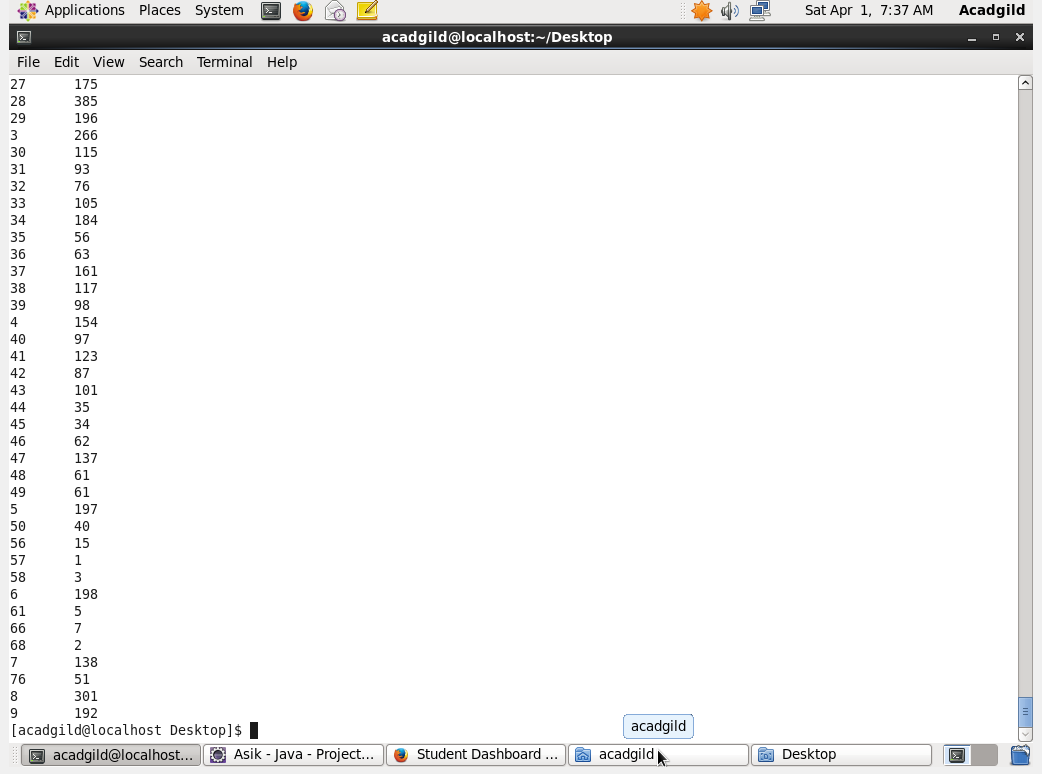
Here as we have the single key all the values related to that key will be added.

Running the Jar File:

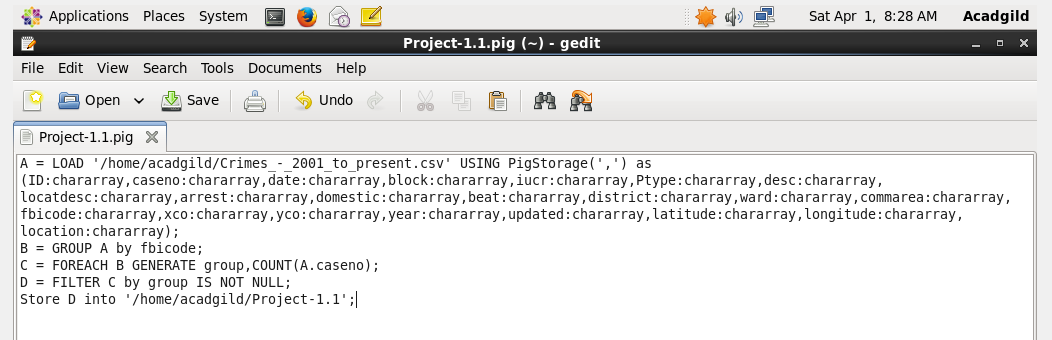


Output:



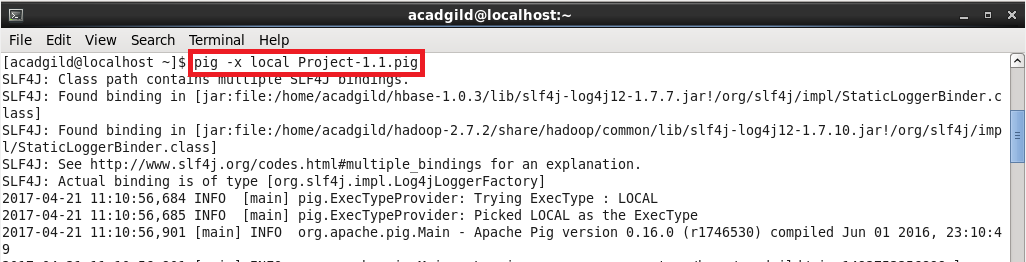


Pig Program:



In the pig script first loading the data in the comma delimited format then grouped by the FBI code and counted the number of codes in each group which gives us the number of cases under each code.

Running the PigScript:



Pig Output:

