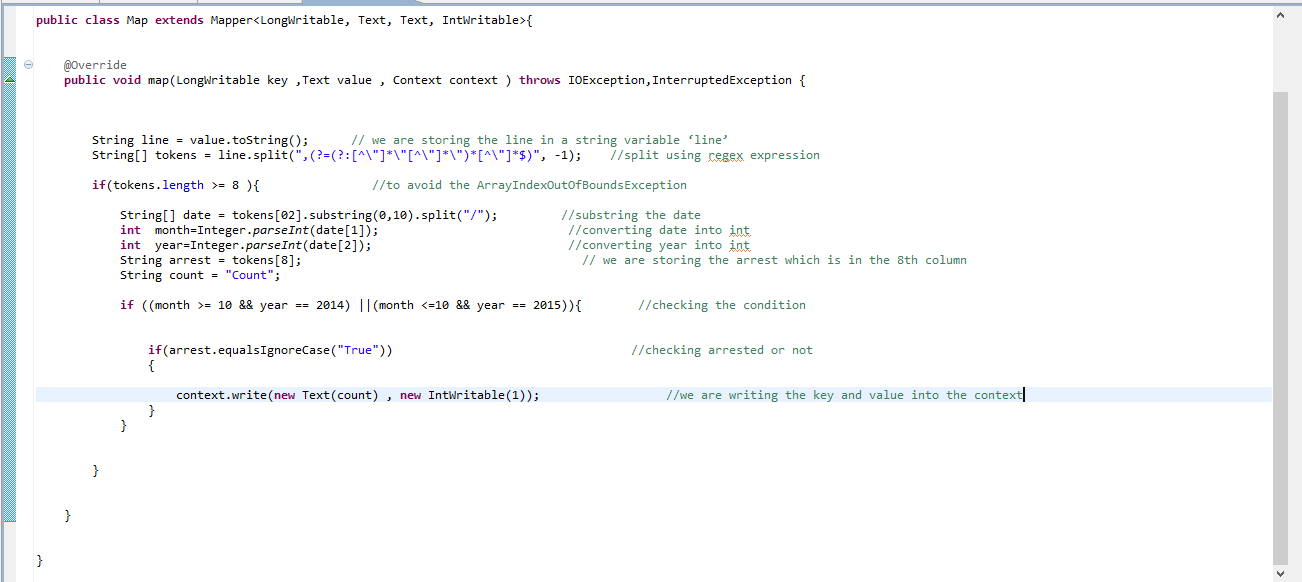
Task 4: Write a MapReduce/Pig program to calculate the number of arrests done between October 2014 and October 2015.

Mapper class:



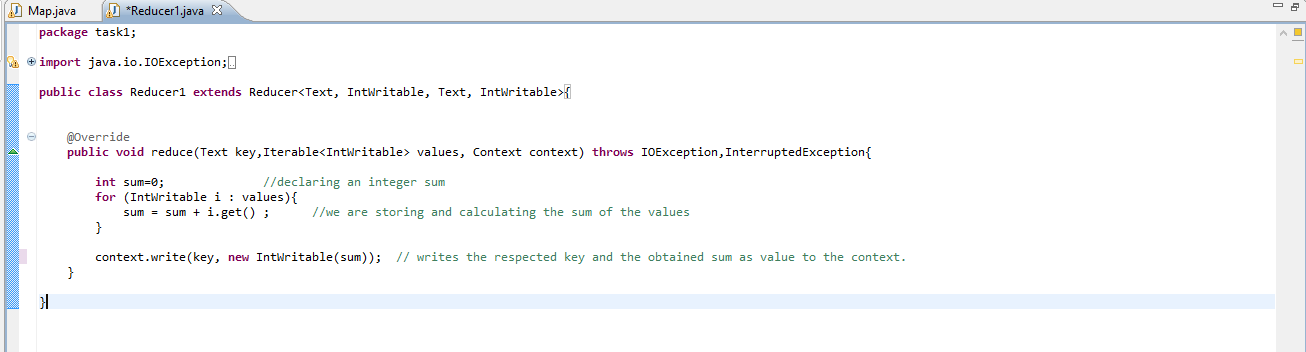
We are splitting the line by using “,(?=(?:[^\"]\*\"[^\"]\*\")\*[^\"]\*$)” delimiter and storing the values in a String Array so that all the columns in a row are stored in the string array. Then we are taking a condition if we have the string array length greater than 8 which means if the line or row has at least 8 columns then it will enter into the if condition and execute the code to eliminate the **ArrayIndexOutOfBoundsException.**

  We are storing the date, arrest which is in the 2nd , 8th column. We are converting that date, year data into int data by type casting.

If condition is used to filter out the data which lies between “10/2014 to 10/2015 as mentioned in problem statement”.

Inner if condition is used to filter out the arrested data only. We are writing the key and value into the *context* which will be the output of the map method.

Reducer class:

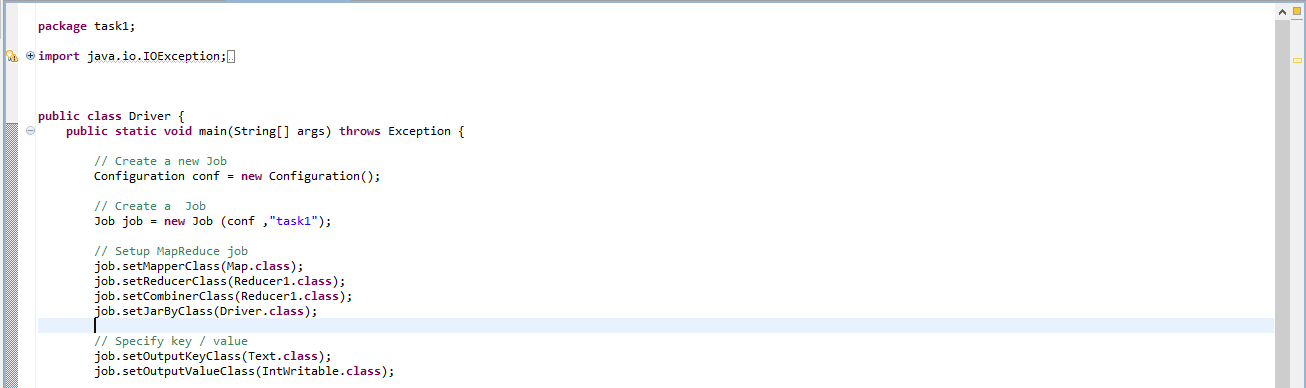


 We are declaring an integer sum which will store the sum of all the ages of people in it. A for each loop is taken which will run each time for the values inside the *“Iterable values”* which are coming from the *shuffle* and *sort* phase after the Mapper phase.

We are storing and calculating the sum of the values.  We are performing the average of the obtained sum and write the respected key and the obtained sum as *value* to the context

Driver class:

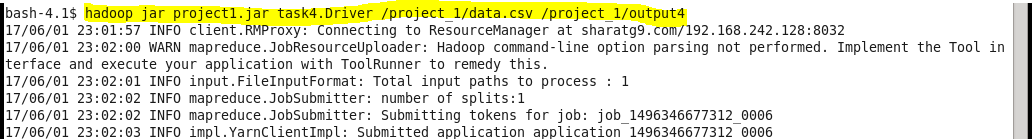
In Driver putting the small file in distributed cache and put the input file and Output file.

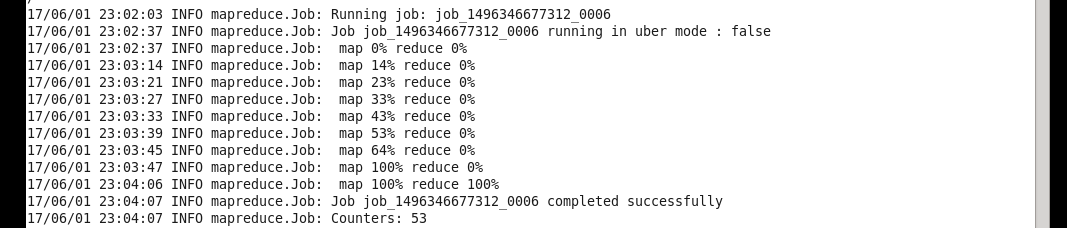




Executing jar file:

hadoop jar project1.jar task4.Driver /project\_1/data.csv /project\_1/output4





Output:

