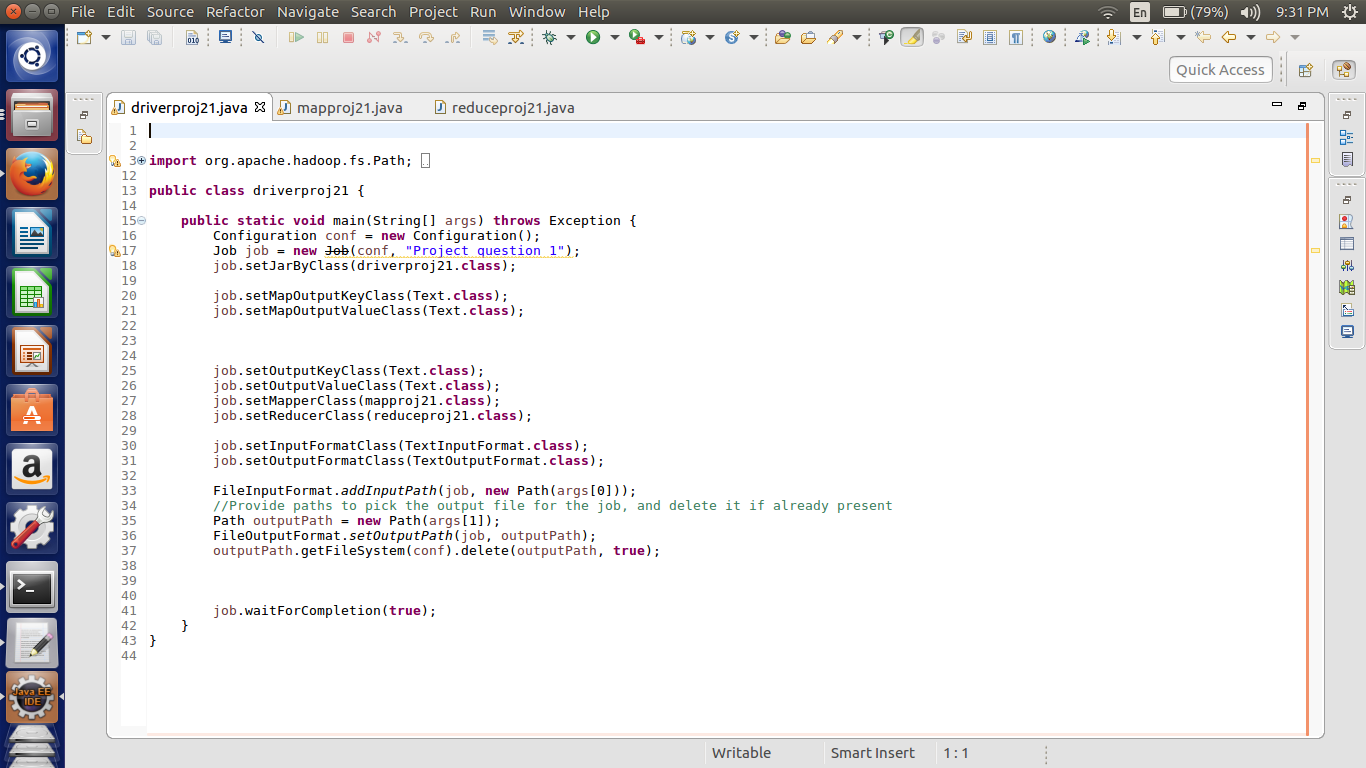
1. Write a mapreduce program to remove commas present inside the double quotes.

// driver class

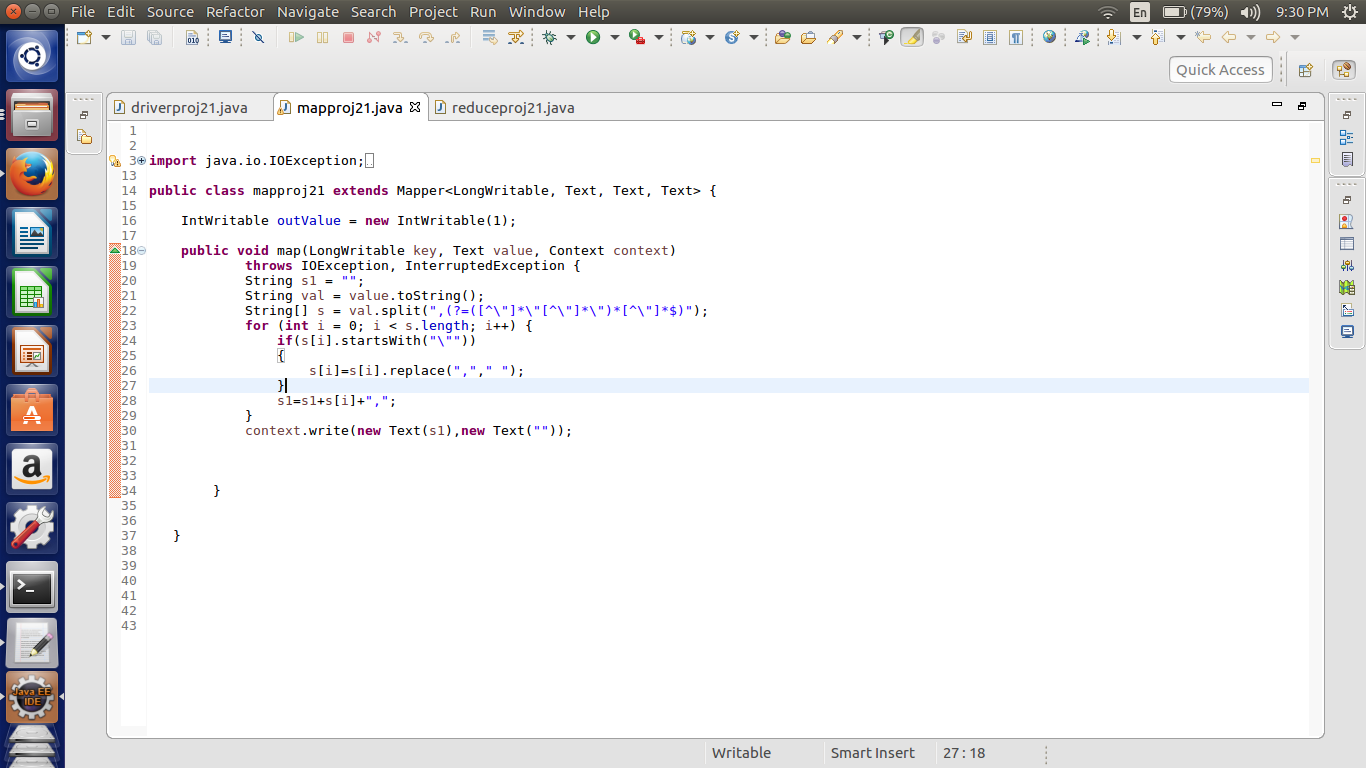


Here we are setting the job configurations and in the first question we can achieve the results by doing the map and reduce job hence setting the in the job configurations about the map class as well as the reducer class.

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.

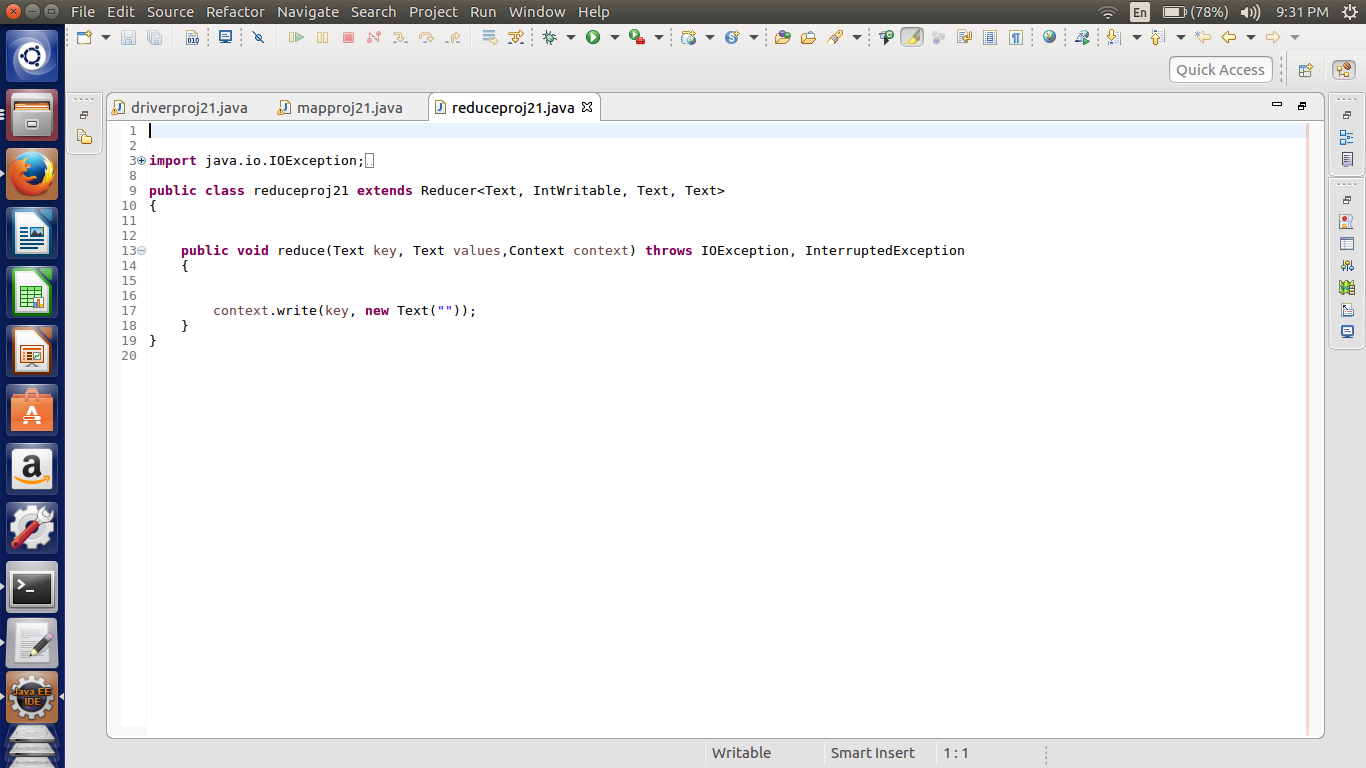
//mapper class



In this mapper class we are splitting the input by using the regex function. The regex function splits the output such that the commas inside the quotes are not considered.

Then after splitting we are finding the segment which starts with quotes and if the starting with quotes then replace the commas in the quotes by blank spaces.

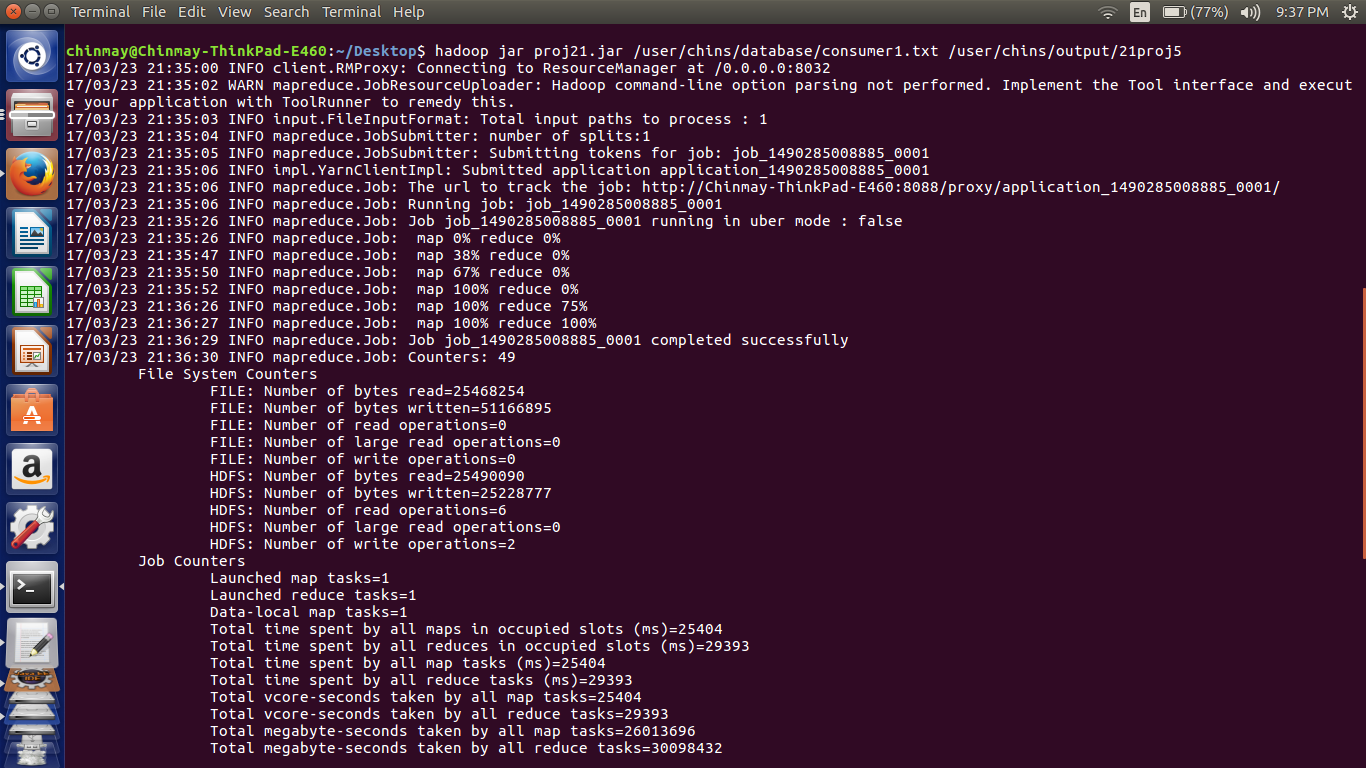
Then again the segments are arranged in the same order as they were and sent to the reducer.

//reducer class

In the above example reducer is doing the default function. Reducer is not necessary for getting the above result.

Map only job can also be done.

//implementation



The output file is attached as consumer sorted. The commas inside the quotes are replace by the space.