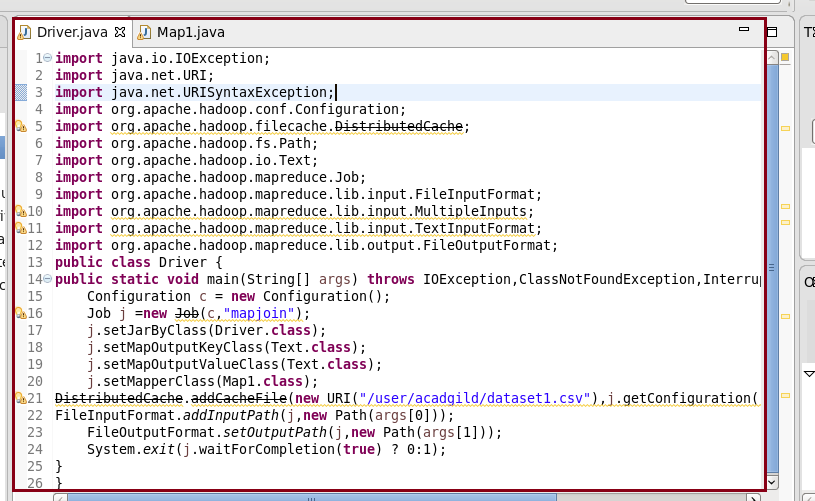
Assignment – 16.5

1. Write a map reduce program to join the below given two datasets using map side join. You can download the datasets from the below link https://drive.google.com/open?id=0ByJLBTmJojjzSWhfXzI5NGh4RGM Dataset description of dataset1.csv is as follows: id,first\_name,last\_name,email,gender,ip\_address Dataset description of tdataset2.csv is as follows: id,first\_name,last\_name

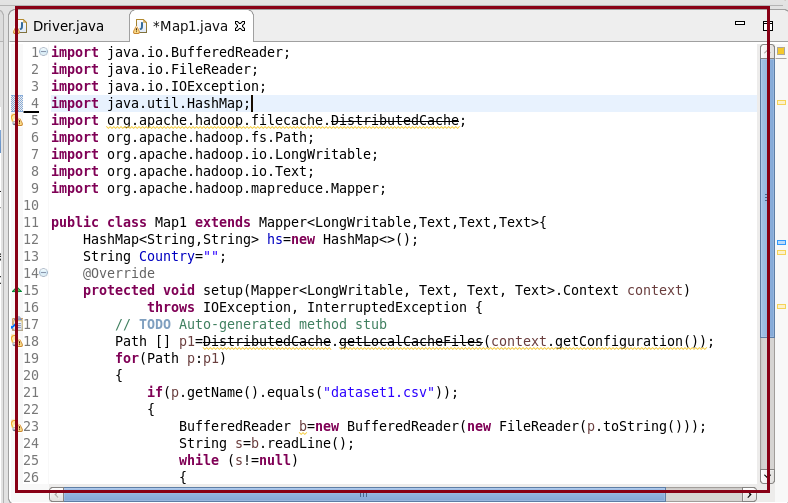
Driver Class:

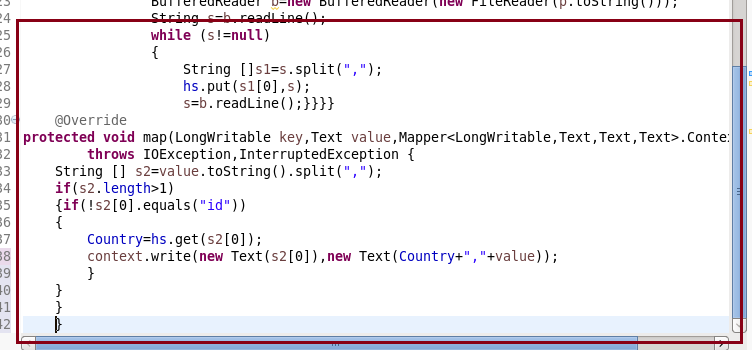


Mapper Class:

In the mapper class we are overriding setup method. It is used to get the dataset from distributed cache and then I used hashmap and storing the id as key of hashmap and total line as value.

In map method I’m sending the id as key and joining both values separated by comma.





**Running jar with output:**

Using cat command I am printing output.

