Query 6

Find the percentage and the count of each case status on total applications for each year. Create a graph depicting the pattern of All the cases over the period of time.

Platform used: Mapreduce

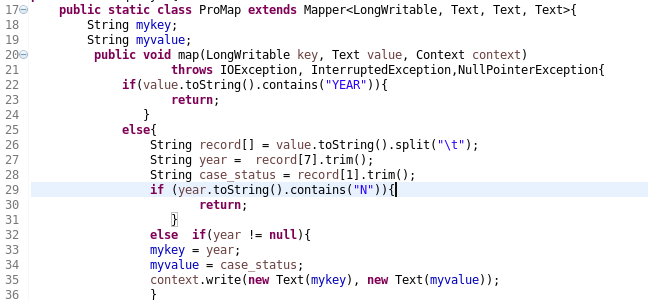
Input file:

/user/hive/warehouse/project.db/h1b\_app2

Explanation:

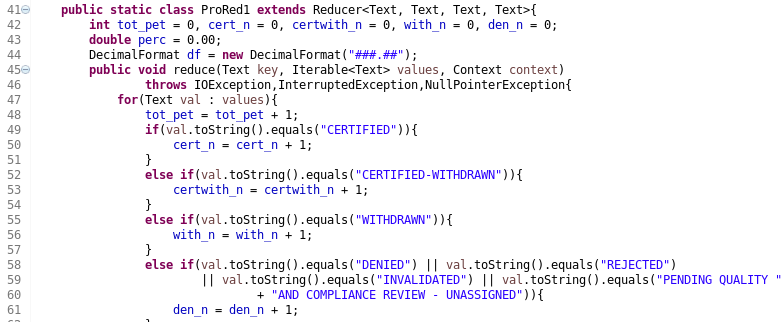
Program – query6.java

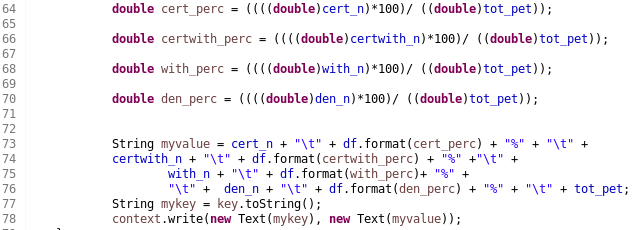
Mapper



In the mapper we have taked year and case\_status as key and value respectively as we want to find the total petitions year wise. We segregate the case statuses later in the reducer.

Reducer





As you can see we have declared separate int variables(cert\_n, certwith\_n etc) for each case statuses and one specific integer variable(tot\_pet) for the total petitions in that year. We start counting petition for each case status and the total as well from lines 47 to 61. Later, from lines 64 to 70 we calculate the percentage of each case status on the total count for each year.

The program is below:



The output file :



The excel file:

