Assignment-9

# Algorithm Optimization

Lines of code : #120

Operation time (execution) : 2 mins

Bottleneck :

There are 449 files we need to run. So those files should be processed to do further analysis and reading all those files at same time is huge bottleneck and making the program execute slower. So I had to implement for loop in the code so that each file is read at a time and add the values to existing list.

Faster implementation :

Data reading and implementation in python is fast but not as fast as when we implement in spark. At the beginning I used python to implement. After that I read the dataset into spark rdd object which made the implementation faster and easy to access.