BIG DATA –Spring 2019

ASSIGNMENT 1

Due Date: 8 Feb 2019 till 10:00 AM on google classroom.

This assignment is just a simple twist to word count problem to ensure that you are capable of executing jobs in map reduce framework over Hadoop.

Question 1. *Write a Map Reduce program in Java for counting long words starting from each English alphabet. A word will be considered long if its length is greater than 5. Ignore the letter case, i.e. consider all words as lower case and ignore words starting with non-alphabetic characters.*

- ***You have to write Map, Reduce and Combine function***

- ***Run your program over the input file provided with tutorial in Standalone mode as well as in the pseudo-distributed mode.***

- ***Upload the Source code and the output file on Google classroom.***

Input:

MapReduce [45] is a programming model for expressing distributed computations on massive amounts of data and an execution framework for large-scale data processing on clusters of commodity servers. It was originally developed by Google.

Output:
A 1
C 3
D 2
E 2
F 1
G 1
L 1
M 2
P 2
S 1
O 1