Big Data Infrastructure and Technologies for Data Analytics

Homework Assignment (HA02)

Note: HA02.01 offers two options: (1) run MapReduce over simple example manually, or (2) use Java program.

For both assignments, you can use explanation and simple example in Lecture Notes 2 (Part 2).

HA02.01. Word count with MapReduce (Individual Hand-In)

You can select one of options: (1) run MapReduce over simple example manually, or (2) use Java program.

For manual option, use text not less than 100 words. It is good to use text with rich language like verses or classical literature :-)
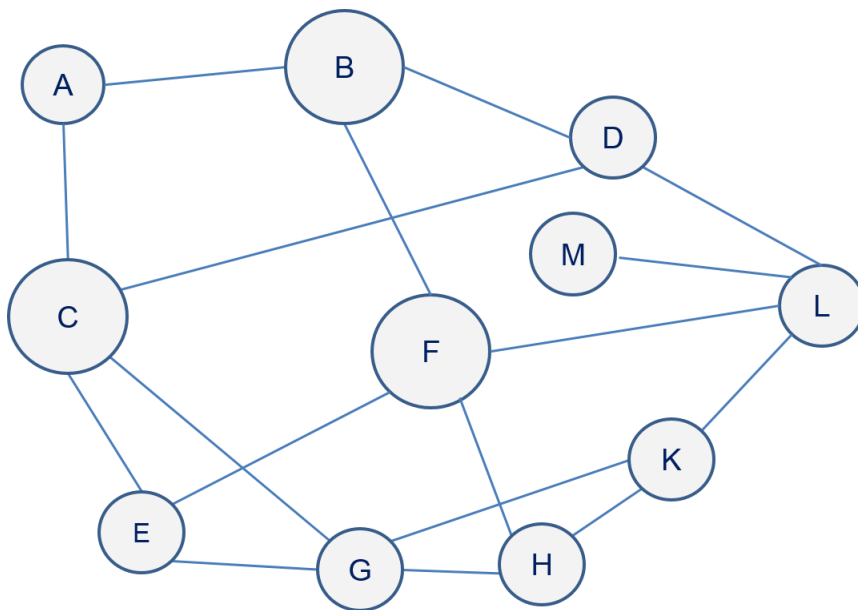
For Java option find a test text of size 100-200K words

You can use one of one of these Java programs to run MapReduce
https://dzone.com/articles/word-count-hello-word-program-in-mapreduce

https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html#Example%3A+WordCount+v1.0

HA02.02. Running Pregel algorithm "manually" (Individual Hand-In)

Take a graph below and select one of nodes as the root node. Calculate the shortest path from the root to all other nodes. Describe the main steps the Pregel algorithm will take and when it converges.

Submission and discussion.

Prepare your answers and submit both answers in one document to designated Assignment link.

Be ready for discussion in the class.