Distributed Systems

Computer Assignment 2 (Kompics)

Due: 22nd of Azar, Midnight

Please upload your code and graphical notation design to the course webpage no later than Friday, 22nd of Azar, midnight.

The goal of this computer assignment is to take as input a text file (the attached text file as an example), and count the number of each word in this file. We want you to implement this part using the MapReduce algorithm. To do that:

1. Use the output of the previous computer assignment and find the node which has the minimum average distance to all other nodes and set it as the root node.

2. Consider each leaf node in the MST as a mapper.

3. Consider each non-leaf node as a reducer.

4. Split the file into smaller parts (number of these parts is equal to number of leaves in the MST), and apply map function in the leaves, and reduce function in the remaining nodes.

Finally, you need to write the output in a file like this:

the:177
need:232
for:354
...

For more information about the MapReduce algorithm, please read the following article: https://www.edureka.co/blog/mapreduce-tutorial/.

Feel free to ask your questions by sending email to: iman.saberi@ut.ac.ir