

The goal of this assignment is to implement and test some modifications to the Map Reduce Application covered in class.

The application should include the following features:

1: The program takes a list of text files to be processed. The file list can be passed to the program via the command line. Use a number of big text or log files for testing your program.

2: Using Approach 3 (multithreaded) in the attached source code modify the given Map Reduce algorithm to build an output data structure that shows how many words in each file begin with each letter of the alphabet e.g. A => (file1.txt, = 2067, file2.txt = 180, ...), B => (file1.txt = 1234, file2.txt = 235, ...) etc then print out the results. The results can also be written out to a file for later analysis etc.

3: Modify the main part of the program to assign the Map or Reduce functions to a Thread Pool with a configurable number of threads. Lookup the Java concurrency utilities for examples of using Thread Pools. The actual number of threads can be passed to the program as a command line parameter.

This assignment can be done either individually or in groups of no more than two students. When completed you should submit copies of the code you have written for the assignment as well as screen shots of the application running. These screen shots should include evidence of having tested your application with some large text files.

If you are submitting for a group then don't forget to supply the full name and ID number of the other group member. All submissions should be done via Blackboard and if you submit more than one attempt then only the final attempt will be marked.