# Distributed System

Assignment 2: Map Reduce

The goal of this assignment is to implement and test some modifications to the Map Reduce Application covered in class. The modified application should be passed 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 large text or log files for testing your program. You can download large text files from the Internet for testing purposes.

1: Modify the main part of the program to assign the Map or Reduce functions (in Approach Three) 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. Run the program and implement some method within the program to measure as accurately as possible how long it takes to process the full set of large input text files used for testing. Does the number of threads in the Thread Pool affect the results?

2: Create another version of the program that also implements parallel processing for the Group phase. As in the previous case, you should use a Thread Pool with a configurable number of threads to handle the parallel processing requirements. Compare the two versions of the program you have implemented by testing and measuring the total time taken to index the same set of large text files.  Which version provides the better performance? How would your modifications to the Group Phase scale if you had a huge dataset to process?

This assignment can be done either individually or in groups of no more than two students. If you are submitting for a group then don't forgot 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. A description of the system implementation and testing, and related screen shots of the applications running, should be included in a single Word or PDF file. These screen shots should include evidence of having tested your application with some large text files. This Word or PDF file should then be included in a standard ZIP file along with all the related Java source code. No other file formats will be accepted for correction. So you should be uploading a single ZIP file for this assignment. Please do not include any of the compiled Java class files in the submitted ZIP file as these are not required.