Omar Badran, Jordan Osecki, and Bill Shaya CS647 Pre-Proposal

Our CS647 group would like to explore the Map and Reduce distributed software system for our term project. We are proposing the development of a Java application to count the number of words in a file. Upon running the application, our software framework will spawn a pre-configured number of worker nodes to simulate a distributed computational environment.

Our group plans to incorporate self adaptation through self healing and self optimization. Self healing will be accomplished by monitoring of worker nodes. If a worker node fails due to loss of connectivity to the network, or some other fatal condition, the failed node’s computation will be redistributed to a healthy node. Therefore the overall computation can seamlessly complete despite the single failure. Our application framework will include a module to incorporate random failures throughout the simulated network in order to exercise self healing. Self optimization will be accomplished by evaluation of the performance of an individual worker node. Our application framework will also include a module to incorporate performance changes in a worker node. As computations are executed, performance will be evaluated, and if necessary, reallocation of computations will be performed in order to optimize computational speed. In order to evaluate the effects of self adaptation, timed metrics will be recorded and analyzed.

There are several notable map and reduce systems that exist such as Skynet and Hadoop. Skynet is an open source Ruby implementation of Google’s MapReduce framework, which is adaptive, fault tolerant, and has only worker nodes which can act as a master at any given time. Hadoop is a Java framework to implement MapReduce functionality, which is currently used in Yahoo web searches.

We feel that our project has adequate scope for a team of three. Work breakdown components will include the master functionality, worker/reducer functionality, self-\* incorporation, and fault detection and handling. Each component can be completed independently by a group member, and we do not anticipate any issues with completing the project by the end of the class term.