

Midterm Status Report

Team Members:

- Pantelis Kanaris
- Giorgos Paphitis

Project Overview

Developing a distributed MapReduce framework in Go, enabling parallel data processing across multiple nodes/machines.

Completed Objectives

- Expanded the lab's implementation for multiple process to multiple nodes.
- Added the functionality for reducers to fetch their assigned intermediate results from the mapper workers through RPC instead of the previous implementation where communication was facilitated through intermediate files
- Separated the master logic from the application logic. Now the master is continuously running listening for incoming jobs from any arbitrary application. The master supports a predetermined set of MapReduce functions and each application selects the function they want, passing the required data.

Next Steps

- Implementing a shared memory space for all nodes, simulating a distributed file system.
- To allow easier scalability, we will also write two Docker images, one for the master and one for the workers.