Saichetan Aitha (sa60)
Sonam Jain (sonyj9)

CS 425 MP1-G12: Distributed Log Query System

Our distributed log query system is built around a client-server model, where each machine can act as both a server (listener) and a client (sender). The server listens for incoming queries from other machines, while the client sends grep queries to each machine. This approach allows us to search log files across multiple machines and collect results efficiently.

We use goroutines to handle incoming connections concurrently. When a server receives a query from a client, it creates a new goroutine to process that query without blocking the server from handling other requests. Each server runs the grep command locally on its machine and then sends the response back to the client which aggregates all results in an output file. We found it more efficient to run grep locally because fetching all the log files to the querying machine would involve significant network overhead, especially for large log files.

We made a test suite that includes test cases to test frequent, infrequent, regex, present in some, and present in one patterns. We created a function that generates log files randomly with these specifications on each machine and then runs the test case. The test suite can run from any machine and will test multiple machines acting as the client.

Grep commands with higher average runtimes correlate to more frequent patterns in the test files. These commands also have larger standard deviations. This is what we expected because these commands have a higher response output which means the system needs to handle more I/O operations and there is more memory usage. There are more factors that can vary from test to test so the standard deviation is higher.