ECE 428 MP1

Ribhav Jain (ribhavj2) and Jiyu Hu (jiyuhu2)

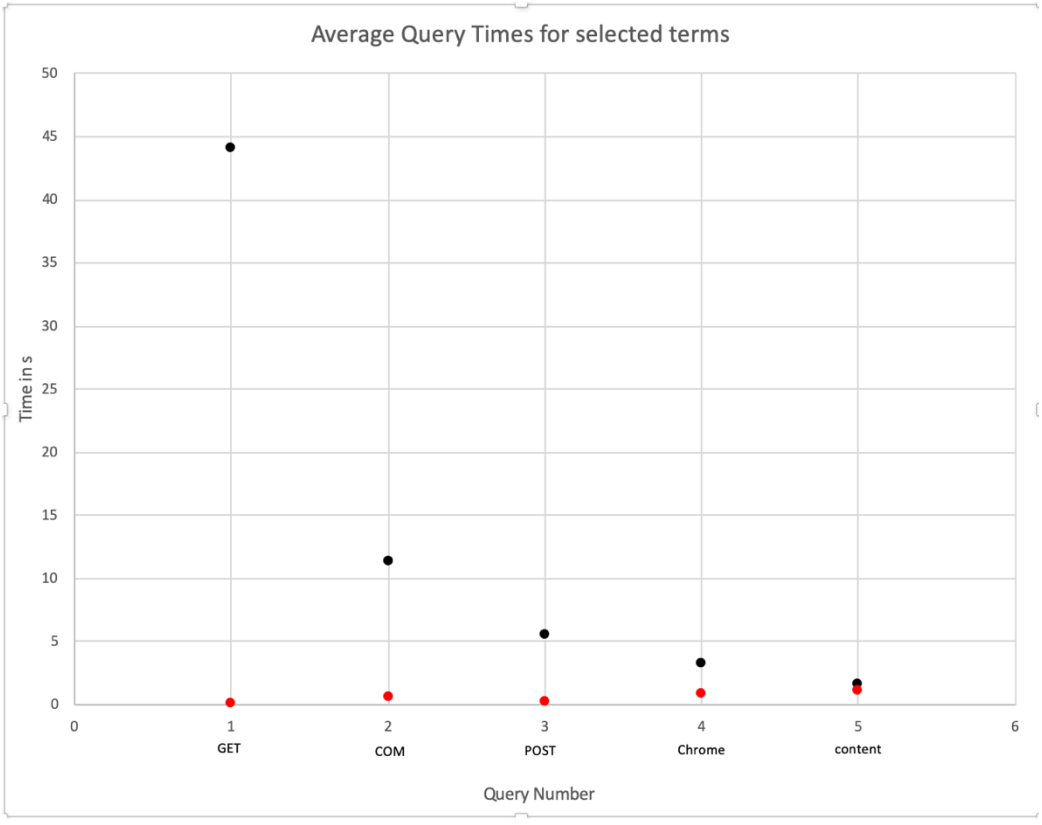
Classes:

1) Multithread: This is a class that is used on the client side to take input words and continue the grep process by calling the Client class on different threads for each VM.
2) Client: This is a class that implements runnable, initiates socket connections to VMs, and receives output from servers.
3) Server: This is a class that is used by the VMs in order to process the grep command. It accepts socket connections, makes grep calls on log files and returns output.
4) UnitTests: This is the class that contains all the unit tests for methods in other classes.
5) testServer: This is a class used for receiving the self created log file for Unit Tests.

Algorithm:

Once the Server code is running on required VMs, the Multithread class is initiated through the terminal with the term to be grepped on the distributed system. The Multithread class will execute multiple client threads simultaneously, each of which will try to connect to a separate Virtual Machine through a socket connection. Each thread will attempt to connect to a VM and give it the term to be grepped. The Server code on receiving the term then greps the respective log file and returns the output to the machine that made the query. The client code only outputs the grep result for a VM when it has returned the entire output for the grep command. The program is fault tolerant and can fetch answers from all machines that have not failed.

Unit tests: We have implemented unit tests to check the successful updating of class variables and states. We have also implemented tests to check socket connections and internet data transfer. We also generate logs on machines and test if they match correctly with what we expect.



Black dots - Average query latency
Red dots - Standard Deviation

Average query latency for common terms when 4 machines each store 50 MB files. We can see the more common terms take longer as they also take longer to print to terminal/file.

We can also see the trend that as the query time decreases there is also an increase in standard deviation.